Designing Keystroke Logging Research in Writing Studies

Mariëlle LEIJTEN Luuk VAN WAES

University of Antwerp, Belgium

Abstract: As typing has become more and more our preferred way of text production, keystroke logging has become one of the major observation tools used in writing process research. It allows for fine grained data collection without intruding into the writers' activities or influencing the writing dynamics. This article describes how keystroke logging — more specifically Inputlog — can be used in writing process research. We elucidate the research flow, starting from the data collection over the data analysis (including pre-processing and merging), and end with some suggestions on how to report findings using different visualizations. We illustrate this research flow by highlighting a few basic analyses that are provided in Inputlog. The article closes with a brief preview on future perspectives on keystroke logging.

Keywords: keystroke logging; writing processes; writing research methods; pause analysis; cognitive processes

1. Introduction

More than ever writing is considered to be one of the most important skills, not only in education, but also in other social contexts. This is mainly due to the rise of mass writing using all kinds of digital devices. Or, as Deborah Brandt phrases it: "For perhaps the first time in the history of mass literacy, writing seems to be eclipsing reading as the literate experience of consequence" (Brandt 2015: 3). Therefore, it is not surprising that the amount of research studies on writing has also increased.

To understand the nature of writing, researchers have used and developed a considerable variety of research methods (Mackey & Gass 2015). Each research method allows researchers to address specific research questions, and all methods have their strengths and weaknesses. Hyland (2016, 2019) offers a comprehensible overview of major research methods in second language (L2) writing. He classifies four broad ways of collecting data related to writing: (1) elicitation: ways of prompting self-report and performance data; (2) introspection: ways of collecting verbal or written reports by text users; (3) observation: direct or recorded data of "live" interactions of writing behavior; (4) text samples: collections of naturally produced samples of writing (117).

In this article, we explicitly focus on the third category, *observation*, and, more specifically on the observation of digital writing processes using keystroke logging. In recent years this way of observing writers has become more and more popular and the possibilities to analyze the resulting logging data have increased rapidly (Lindgren & Sullivan 2019). Keystroke logging comprises a logging program that is activated on a computer allowing the researcher to record every keystroke and mouse click or movement related while a participant is writing. These logging events are time coded so as to exactly reconstruct the writing process and analyze the writing process dynamics in function of time and cognitive effort (Leijten & Van Waes 2013).

Keystroke logging (KSL) mainly focuses on characterizing different aspects of (cognitive) writing processes. In contrast with other methods that, for instance, mainly address aspects of the writing product (e.g., text quality or text complexity), the writer him- or herself (e.g., self-efficacy) or the social context (e.g., collaboration or feedback). We contend that studying the dynamics of the writing process leads to important insights in writing and a better understanding of the complexity of writing, complementary to other research methods. Moreover, one of the advantages of keystroke logging is that it easily allows to combine observation with product related research methods.

A metaphor we sometimes use to explain the importance of studying processes compares writing with cooking. When you go to a restaurant, and you eat a delicious dish, it is not guaranteed that you will be able to cook that dish yourself. However, when the cook allows you to watch how he prepares the dish, that will certainly enhance the chance that you will be able to prepare the dish yourself, mainly because you have been able to observe the preparation and cooking process. That explains why, at least in Europe, cooking programs on television are so popular. You learn new techniques, ingredients, tips, and tricks that are hard to induce from a finished product. The same holds for writing products and processes.

When designing keystroke logging studies, it is important to have knowledge of the theoretical framework that this type of writing process research draws upon (Galbraith & Baaijen 2019). The well-known process models offer a solid basis to set up and interpret writing process studies. Moreover, recent studies also aim at further building a solid theoretical foundation (Conijn, et al. 2019). In the latest revision of Hayes' model, Leijten, Van Waes, Schriver and Hayes (2014) have tried to represent how (professional) writers produce texts based on an empirical keystroke logging study (see Figure 1).

The writing model in Figure 1 builds on the earlier models by Hayes (Flower & Hayes 1981; Hayes 1996, 2012) and accounts for the processes that individual writers engage in as they plan, compose, and evaluate their texts. While focusing mainly on cognitive processes, the model considers some aspects of social processes as well. In brief, the model has three levels: (1) a control level, (2) a process level, and (3) a resource level. The *control level* includes motivation and the processes and structures that control the other writing processes. The *process level* has two major parts: on the one hand, the "writing processes" representing the internal mental processes that the writer uses to compose; on the other hand, the "task environment" writers interact with during the process (i.e., the physical, social, and cultural contexts of the writing processes). Finally, the *resource level* includes long-term memory, working memory, reading, and the ability to focus attention.

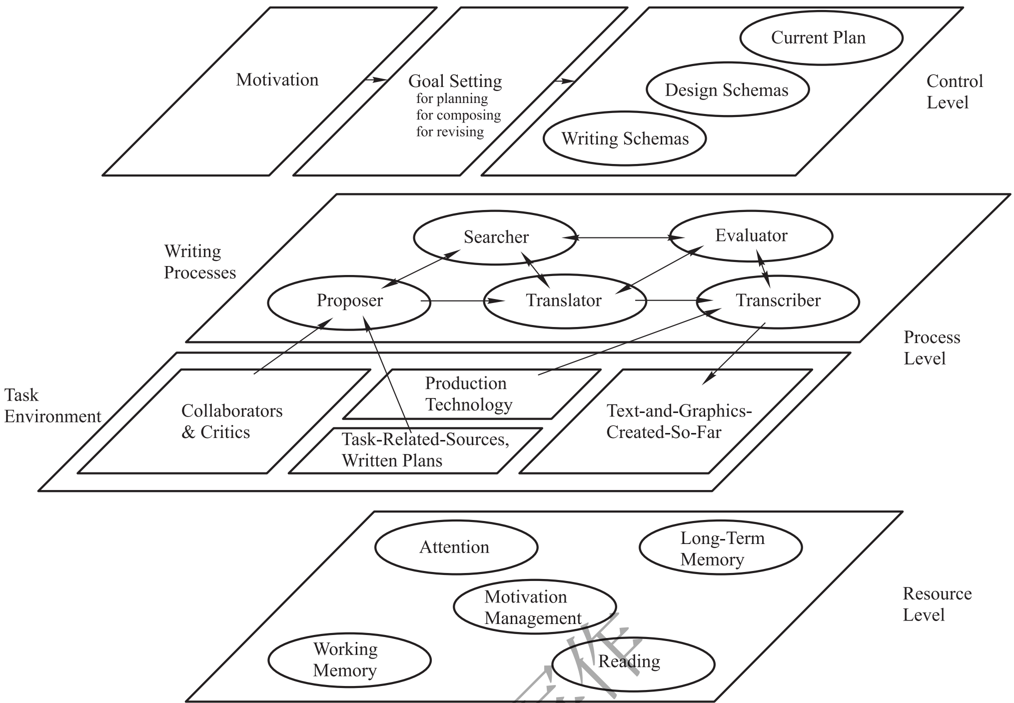

Figure 1. Model of composing elaborated to encompass activities of skilled professional communicators (Leijten, Van Waes, Schriver & Hayes 2014)

In this article we will illustrate how keystroke logging can be used to further explore, understand and test the interaction of the (cognitive) processes at different levels, as well as the interaction between the levels. Currently three European logging tools are widely used for writing and translation research: Scriptlog (Sweden; Johansson et al. 2014), Translog (Denmark; Carl 2012) and Inputlog (Belgium; Leijten & Van Waes 2013). These three logging tools all have their own specificities making them suitable for different types of research studies. In short, *Scriptlog* is most suitable for (highly) controlled research designs, e.g., it has an in-house text editor which can be combined with visual input. The program also allows the integration of SMI eyetracking. *Translog* is mostly used in translation studies that focus on the writing and reading processes of source and target texts in different languages. It is also connected to SMI eyetracking tools. *Inputlog* is mostly suitable for studies in which writers are working in a Windows environment, using Microsoft Word and other Windows based programs. Inputlog has been initiated in 2003, as a counterpart for other logging tools that were mainly designed for experimental research studies. It is a tool designed to log and analyze writing processes in both ecological and experimental settings. In this article we focus on designing keystroke logging research based on Inputlog. Typical for Inputlog is that it logs in whatever Windows environment and relates these events to a time stamp for the start and end of each event (in milliseconds). When writing in MS Word extra characteristics that relate to the input events are logged to allow for refined writing analyses: position (doc position of the character or x-y coordinates of the mouse click), actual document length (number of characters produced at each moment in time), and copy-paste actions. The program also logs text production with speech recognition (Dragon Naturally Speaking, Nuance) and tracks activities that relate to the use of external digital sources (e.g., other documents or URLs on the Internet). Finally, keystroke logging data can be merged with Tobii eyetracking data (as provided by Tobii Studio).

We will start this article with a brief review of the kind of research themes that are currently investigated using Inputlog. As the choice of a research method involves both theoretical and practical considerations, in the next section we describe the typical research flow that characterizes the design of writing process studies using Inputlog. The steps in the research process are illustrated for the use of Inputlog as a research tool (see www.inputlog.net; Leijten & Van Waes 2013). Next, we focus on some exemplary data analyses showing that each analysis provides a specific perspective on the collected logdata. We conclude this article by looking ahead at further developments in keystroke logging in general, and more specifically, in Inputlog.

2. Themes in Writing Process Research

In recent years numerous research studies have focused on writing process research. In their latest book on observing writing with logging tools, Lindgren and Sullivan (2019) report nearly 200 studies in their general introduction. In this article we would like to narrow down the overview to studies that refer to Inputlog. We have selected three general papers on the use of Inputlog (Leijten & Van Waes 2006, 2013; Van Waes & Leijten 2006) and report on the main English and Chinese contributions since 2010 that refer to these three articles. We used Google Scholar, Baidu, and CNKI for this search. In total 491 studies refer to these three papers, of which about 364 are unique references. The current overview is based on English articles, books, and book chapters (338 publications). In the first step we categorized the studies via NVivo in three main categories: contributions that focus on *methodological issues* related to keystroke logging research, *theoretical papers* and *applied research* studies (Table 1). More than fifty percent (57%) of the papers present applied research. However, as keystroke logging still is a fairly new research method, methodological papers take a fair — and gradually increasing — share (32.8%). As expected, fewer papers take a more fundamental perspective and address theory building (9.4%).

Table 1. Distribution of 338 publications that refer to three main references of Inputlog

	Methodology (n=111)	**Theory Building (n=32)**	**Applied Research (n=194)**
2014 and previous	31	15	54
2015	10	2	20
2016	13	6	24
2017	12	3	32
2018	17	5	19
2019	28	5	35
2020 (until April 2020)	3	1	10

In a second step we categorized all titles in relevant subthemes to provide a more detailed and elaborate overview of research topics in the various fields of writing process research (Table 2).

Table 2. Overview of a selection of KSL publications between 2010-2020
(full reference list available on http://www.inputlog.net/downloads)

Theme	**Articles**
Methodology	General advice on measuring cognitive writing skills Benetos & Bétrancourt 2015; Galbraith & Baaijen 2019; Hyland 2016; Medimorec & Risko 2017; Van Waes & Leijten 2015; Van Waes, Leijten, Lindgren & Wengelin 2015; Zhang, Hao, Li & Deane 2016; Zhu, Zhang & Deane 2019
	Low level writing processes (typing measures) Aldridge & Fontaine 2019; Conijn, Van Zaanen, Leijten & Van Waes 2019; Guo, Deane, Van Rijn, Zhang & Bennett 2018; Medimorec, Young & Risko 2017; Van Waes, Leijten, Mariën & Engelborghs 2017; Van Waes, Leijten, Pauwaert & Van Horenbeeck 2019
	High level writing processes (linguistic analysis) Cislaru & Olive 2018; Leijten, Van Horenbeeck & Van Waes 2019; Leijten, Van Waes & Van Horenbeeck 2015; Medimorec et al. 2017
	Mixed methods (combining Inputlog with protocols, eyetracking and/or additional tools) Chukharev-Hudilainen, Saricaoglu, Torrance & Feng 2019; De Smet, Leijten & Van Waes 2018; Hyland 2016a; Revesz, Michel, Lu, Kourtali & Borges 2020; Wengelin, Frid, Johansson & Johansson 2019

(to be continued)

(continued)

Theme	**Articles**
General theory building	Writing development Min 2017; Rogiers, Merchie, De Smedt, De Backer & Van 2020
	Writing models Baaijen & Galbraith 2018; Leijten, Van Waes, Schriver & Hayes 2014; Olive 2014; Van den Bergh, Rijlaarsdam & Van Steendam 2016; Van Waes & Leijten 2015
	Theory of translation Dam-Jensen, Heine & Schrijver 2019; Doherty 2016; Ehrensberger-Dow & Massey 2014
Applied research	Translation and computer-aided translation Bundgaard, Christensen & Schjoldager 2016; Carl, Schaeffer & Bangalore 2016; Daems & Macken 2019; Ehrensberger-Dow & Perrin 2013; Hanoulle, Hoste & Remael 2015; Robert 2014; Schrijver, Van Vaerenbergh, Leijten & Van Waes 2016; Teixeira & O'Brien 2017; Zapata 2016
	L1-L2 research Barkaoui 2016, 2019; Cho 2018; Choi 2016; Xu, 2017; Michel 2017; Ranalli, Feng & Chukharev-Hudilainen 2018; Révész, Kourtali & Mazgutova 2017; Xu 2018; Xu & Xia 2019; Yuguo, 2019; 袁辉 & 徐剑 2014
	Educational research De Smedt et al. 2018; De Smet, Brand-Gruwel, Leijten & Kirschner 2014; Guo, Zhang, Deane & Bennett 2019; Kim 2020; Nie 2014; Van der Steen, Samuelson & Thomson 2017; Van Waes, Van Weijen & Leijten 2014; Zarrabi & Bozorgian 2020; Zhang, Bennett, Deane & Van Rijn 2019
	Professional writing and translation Bundgaard 2017; Bundgaard & Christensen 2019; Daems, Vandepitte, Hartsuiker & Macken 2017; Leijten et al. 2014; Robert & Brunette 2016; Schrijver, Van Vaerenbergh, Leijten & Van Waes 2017; Vandendaele, De Cuypere & Van Praet 2015
	Writing difficulties Antonsson et al. 2018; Beers, Berninger, Mickail & Abbott 2018; Johansson-Malmeling, Hartelius, Wengelin & Henriksson 2020; Richards, Abbott & Berninger 2016; Van Waes et al. 2017
	Source-based writing Chan 2017; Choi 2016; Leijten, Van Waes, Schrijver, Bernolet & Vangehuchten 2019; Leijten et al. 2014; Revesz et al. 2020; Vandermeulen, Van den Broek, Van Steendam & Rijlaarsdam 2020

2.1 Papers focusing on methodology

Although keystroke logging is a widely used tool in writing research, many authors report on the challenges that still remain to relate logs to underlying cognitive processes (Galbraith & Baaijen 2019; Lindgren & Sullivan 2019). By design, keystroke logging data always requires (elaborated) inferences to interpret and make sense of the data collected.

A first group of papers provides techniques on defining and extracting general measures of cognitive writing skills. Others focus more specifically on low level writing processes like typing skills, or high-level writing processes like syntactic and lexical processing. However, the common denominator in these articles is that the researchers suggest frameworks that relate the raw data of Inputlog to high and low levels of cognitive processes involved in writing. Recent papers also clearly describe advanced statistical approaches and illustrate a more differentiated approach to concepts like pausing and fluency. A separate group of papers are methodological papers describing the combination of Inputlog with other observation tools and techniques in a mixed methods approach, for instance, using both keystroke logging and think-aloud protocols or (stimulated) retrospective interviews.

2.2 Papers focusing on theory building

Not all papers that refer to Inputlog are actually using Inputlog or one of the other keystroke logging tools, but they are more specifically concerned with theory building of cognitive writing processes. However, in these papers the necessity of approaching a more detailed and functional account of writing is elaborated. An approach made "possible thanks to the increasing use of online sophisticated paradigms in writing research which [...] underline the necessity for online studies of writing" (Olive 2014: 189). Other papers make use of experimental research via Inputlog for theory building as we described in the introduction (Leijten et al. 2014). In general, three main themes are prominent: writing development, writing models, and theory of translation.

2.3 Papers describing applied research

Keystroke logging as an observation method has also been widely adopted in translation research (i.e., translation from one language into another language). Translation is an area of professional communication that is closely related to writing process research (Schrijver et al. 2014). Both disciplines study similar concepts (e.g., writing from multiple sources, revision procedures), and use similar methodologies. Although Translog has been specifically developed for experimental translation studies, some researchers who are interested in translation processes in professional settings, opt for using Inputlog. In addition, because of the increasing importance of computer-aided translation, post-editing is becoming a more prominent research topic. The technical possibilities of merging/aligning keystroke logging data with machine translation software packages like Trados and Casmacat create innovative perspectives in this research area.

Inputlog 8 enables researchers to log writing processes in a wide range of languages using Western characters. Therefore, Inputlog is also used in all kinds of second language research. In China it is also suitable for studies in the field of English as a foreign language (EFL), which is one of the most prominent fields in language teaching and learning. Inputlog is now also suitable to log Chinese characters (Pinyin input; cf. Section 5 for more information on Inputlog 9-Beta), opening perspectives for Chinese based L1-L2 studies.

In educational research, various studies are focusing on the interplay between writing processes and text quality. Various influences of process characteristics like pausing patterns, writing fluency, or process approaches like constructing an outline are related to text quality. Various researchers also focus on writing assessment (e.g., ETS in the US).

As stated before, Inputlog is especially developed to be used in professional settings. It is quite difficult to specify what makes a writer a professional or an expert. Several studies have started to profile some of the cognitive processes professionals engage in as they work in various organizational settings. This holds also for other areas of expertise such as journalism and the before mentioned translation studies.

Inputlog is also used in studies to describe writing difficulties, e.g., dyslexia or aphasia. These studies mainly focus on low-level processes, because the writing processes can be highly disruptive. Patients with low-grade glioma and people with Alzheimer's disease are also subject of study.

Finally, recent studies in the field of source-based writing and other types of reading-to-write or integrated writing studies are using Inputlog. Since 2015 Inputlog has enabled researchers to code and categorize digital sources.

3. Designing Keystroke Logging Research: The Research Flow

When one wants to conduct writing research with keystroke logging, it is important to carefully design the research flow. In this section we want to describe a step-by-step procedure we would like to propose as a guide for designing rigorous keystroke logging research in the domain of writing studies (see Figure 2). To make this research procedure concrete, we show how the keystroke logging program Inputlog supports each of the design stages.

INPUTLOG MODULES

Inputlog features five modules that systematically guide researchers through the keystroke logging research process:

1. *Record*: This module logs (keyboard, mouse, speech, and window change) data in Microsoft Word and other Windows based programs together with a unique time stamp (ms).

2. *Pre-process*: As it is often necessary to refine logged data prior to analysis, this module allows researchers to process or filter data from various perspectives: event based (keyboard, mouse, and speech), time based or based on window changes (sources: MS Word, Internet etc.). For instance, the filter is convenient for deleting logging session start-up or deactivation “noise”. For example, when additional questions are asked after the logging session has already been started, this pausing time (noise) can be excluded from the data analyses post-hoc.

3. *Analyze*: This module is the heart of the program and features three basic process representations (general and linear logging file and the s-notation of the text) and four aggregated analyses (summary, pause, revision and source analyses). Additionally, a process and fluency graph is produced. In the current version, a linguistic process analysis is also offered for English and Dutch, which returns the results from a part-of-speech tagger, a lemmatizer, a chunker, the syllable boundaries, and word frequencies.

4. *Post-process*: This module integrates log files from Inputlog with files from other observation tools (Morae, Dragon Naturally Speaking, Eyetracking data). It is also possible to merge multiple output files from a group of participants for group or bulk (statistical) analysis in, for instance, Excel, SPSS or R.

5. *Play*: This module allows researchers to play back the recorded session at various levels (time or revision based). The replay is data based (not video based) and the play speed is adjustable. A logged session can also be reconstructed revision by revision.
(Note: In the current version this module is not fully reliable. Therefore, if researchers’ research is dependent on a process replay, we recommend that they combine the logging with a screen recorder.)

Inputlog is freely available for non-commercial use by researchers in the context of writing and translation research. It is available on: www.inputlog.net. A more detailed description of the program is presented in Leijten and Van Waes (2013).

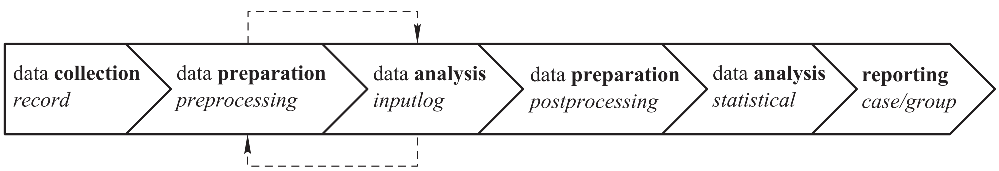

Figure 2. Keystroke logging research flow

3.1 Data collection

After having defined the research questions or hypotheses and having carefully designed the writing tasks, one can start with their writing study. In most cases the researcher will start with data collection (see Figure 2). One of the main decisions that impact data collection is the choice for either an experimental or a more ecological environment. Inputlog facilitates data collection in both contexts, and it is up to the researcher to carefully balance the pros and cons of both options. Moreover, one of the main advantages of keystroke logging is that it is quite unobtrusive and therefore allows for a mixed methods approach or triangulation. Many researchers have, for instance, combined keystroke logging observation with think-aloud protocols, screen recording or eye tracking (e.g., Lindgren et al. 2011; Schrijver et al. 2016; Wengelin et al. 2009).

Before starting a study, we always recommend to carefully check and test the logging environment by conducting a pre-test during which researchers can also pilot their research script, instructions, and the technicalities related to it. As computer configurations and security settings often (slightly) deviate from the expected default setting, it is very recommendable in that stage to check the logging data and verify whether the log files include all the variables the researchers have targeted as indicators to operationalize their research question (e.g., character position, source identification). The output of the general analysis is most suitable for such a check.

The Inputlog Record-module interface guides researchers through the different substeps. After completing the session identification (describing participant and task characteristics), Inputlog allows for different logging options: researchers can either start logging in an empty (MS Word) document, or they can choose to continue working in a previous version of their text (e.g., multi-drafting), or even open another document (e.g., a text by someone else that needs revision).

Recently we also added a so-called “copy task” to Inputlog 8 (Van Waes et al. 2019). This task measures the writers’ typing skills at different lexical levels and allows researchers to take into account differences in typing skills between writers when analyzing log data. We contend that measuring a personal interkey interval baseline is important when comparing, for instance, pause and fluency data among participants. Therefore, we strongly recommend using this task prior to any other data collection. It only takes five minutes and is currently available in eleven languages.

When one starts a logging session, all the log data are stored continuously in an XML-data file. This file is completed with product information of the final text upon ending the log session. Moreover, at that moment the final document is automatically stored locally. The setting menu allows one to define the working directory for the data storage. All data are now collected and ready for further processing.

3.2 Data preparation: Preprocessing

No matter how carefully researchers have collected their data, in most cases it is necessary to evaluate and further prepare the raw data set before analyzing these data. The Inputlog “Preprocess”-module provides some tools to do this at different levels. Usually the data preparation starts with a careful inspection of the general file (see output example in Table 3).

Table 3. Excerpt from a general analysis in Inputlog

#ID	Event Type	Output	Position	Document Length	Character Production	Start Time (ms)	Start Clock	End Time (ms)	End Clock	Action Time	Pause Time	Pause Location
0	mouse	Movement	0	0	0	15	0:00:00	1123	0:00:01	1108	15	INITIAL
1	focus	TASKBAR	0	0	0	1201	0:00:01	1201	0:00:01	0	0	CHANGE
2	mouse	LEFT Click	0	0	0	1201	0:00:01	1279	0:00:01	78	0	INITIAL
3	mouse	Movement	0	0	0	1326	0:00:01	1716	0:00:01	390	47	MOUSE
4	mouse	LEFT Click	0	0	0	1996	0:00:01	2059	0:00:02	63	280	MOUSE
5	focus	WordLog	0	0	0	2340	0:00:02	2340	0:00:02	0	0	CHANGE
6	mouse	Movement	0	0	0	2340	0:00:02	3525	0:00:03	1185	0	MOUSE
7	keyboard	LSHIFT	0	1	0	118597	0:01:58	118956	0:01:58	359	115072	COMBINATION KEY
8	keyboard	T	0	1	1	118844	0:01:58	118900	0:01:58	56	247	BEFORE SENTENCES
9	keyboard	h	1	2	2	119468	0:01:59	119532	0:01:59	64	624	WITHIN WORDS
10	keyboard	i	2	3	3	119796	0:01:59	119876	0:01:59	80	328	WITHIN WORDS
11	keyboard	s	3	4	4	119972	0:01:59	120052	0:02:00	80	176	WITHIN WORDS
12	keyboard	SPACE	4	5	5	120108	0:02:00	120196	0:02:00	88	136	AFTER WORDS
13	keyboard	i	5	6	6	121092	0:02:01	121156	0:02:01	64	984	BEFORE WORDS
14	keyboard	s	6	7	7	121204	0:02:01	121268	0:02:01	64	112	WITHIN WORDS
15	keyboard	BACK	7	8	7	122388	0:02:02	122444	0:02:02	56	1184	REVISION
16	keyboard	BACK	6	7	7	122772	0:02:02	122836	0:02:02	64	384	REVISION
17	keyboard	w	5	6	8	123389	0:02:03	123460	0:02:03	71	617	WITHIN WORDS
18	keyboard	a	6	7	9	123564	0:02:03	123636	0:02:03	72	175	WITHIN WORDS
19	keyboard	s	7	8	10	123732	0:02:03	123804	0:02:03	72	168	WITHIN WORDS
…												

Table 3 shows the logged data at the lowest level. It is a linear, event-based, vertical representation of the text production that is generated by the General Analysis. It shows the characteristics of each logging event (input) in a separate row. In the output column, for instance, the contents of each keystroke: in events 7 through 11 the word "This" is typed. In the position column the numbers go up until a typing error is corrected (event 15-16) showing the recursivity at that instance. Columns 8-11 show the timing information related to each start and end of an event, both in milliseconds and clock time (e.g., pushing and releasing a key). The action and pause time in the next columns are derived from this information. The last column is an algorithmic and hierarchical identification of each event level (e.g., before sentence, within words).

The example in this output indicates that the first letter typed in this document is the (capital) letter T (Shift + t; event 7-8). The starting time that relates to these events is about two minutes after starting the program. Depending on the instruction, the researcher will have to decide whether these two minutes are part of the writing process (e.g., initial planning) or whether it should be considered as noise (e.g., because the researcher provided instruction during these two minutes). In the latter case, the preprocessing time filter allows the researcher to ignore this time period as the filter creates a new idfx log file that can be used in the follow-up analyses. Apart from the time filter to remove logging noise in the beginning and the end of the writing session, Inputlog also provides filters to isolate certain working environments (e.g., only typing in MS Word, and not in Baidu) or input events (e.g., only keyboard actions, and no mouse clicks).

Another preprocessing option refers to writing from multiple sources. Inputlog identifies all window changes during writing, e.g., when the writers leave MS Word to consult an online dictionary or check their email. The source recorder makes it possible to group and classify the consulted sources in functional categories for further analyses (e.g., one can classify a range of online dictionaries in one category called "dictionaries"). This makes it easier to compare writing processes of different writers at a higher and more conceptual level (Leijten et al. 2019). Finally, preprocessing tools are offered to segment or merge logged files that were collected in consecutive sessions. Finally, it is also possible to merge Inputlog files with eyetracking log files (i.e., Tobii) or dictation software (i.e., Dragon Naturally Speaking, Nuance).

3.3 Data analysis in Inputlog

The next step in the research flow comprises the data analysis in Inputlog (see Figure 2). Currently sixteen different analyses are presented in the analysis module, offering a wide variety of options to investigate the log data from diverging perspectives. Table 4 provides an overview of the analyses by classifying them in three categories: basic, specific and visual analyses.

Table 4. Overview of Inputlog analyses

Basic analyses	**Specific analyses**	**Visual analyses**
general	source	process graph
linear	fluency	source network
summary	bigram	fluency graph
pause	word pause	
revision	linguistic	
S-notation	token	
	copy task	

The output of the analyses is both suitable as a basis for qualitative and quantitative analysis. As the current article does not allow to go into detail about all the possibilities, we opted to briefly discuss a selection of analyses in Section 4. For a more detailed description of all the analyses, see the Inputlog manual (Leijten & Van Waes 2019), Leijten and Van Waes (2013) and Leijten et al. (2019).

3.4 Data preparation: Postprocessing

When a research study aims at describing a specific (set of) case studies, the output files do not need further processing. All output files are generated as graphic files or XML files linked to style sheets that can be opened in an Internet browser (preferably Internet Explorer) and can be copied or exported to other programs. However, if a study targets statistical analyses at the group or condition level, the researchers will need to further prepare their data to make them suitable for statistical programs like SPSS, R or MLWin. The postprocessing module in Inputlog facilitates this part of the research flow.

The main option in this module allows researchers to merge (large) sets of data analyses into one file. The merging process is a time saving process and is used to combine analysis files from several participants — or consecutively logged sessions of one participant — into one combined file. For instance, if researchers have logged the writing process of a group of 50 students and generated a summary analysis for all of them, they can merge these analyses and will get one spreadsheet comprising the participants in the columns and the output variables in the rows. The resulting file is a comma separated file (csv-format) that can be opened and processed in MS Excel or any statistical program. A manual check of the final data set is recommended to avoid technical incompatibilities.

3.5 Data reporting

A final step in the research flow concerns the data reporting (see Figure 2). When describing keystroke logging research, it is important to report the decisions the researchers made in the different stages of the research flow in detail to provide the reader with enough information to evaluate and replicate the study. We recommend to always include technical information that refers to: logging program used (and key reference of the program), version used in the phase of logging and analysis, logging conditions and task instruction, type of preprocessing, analyses executed, pause threshold used in the analyses etc. Also consider the possibilities of combining quantitative and qualitative report perspectives as these perspectives might complement each other.

4. Perspectives to Keystroke Logging Data Analysis

As mentioned above, Inputlog provides a large set of analyses. In this section we highlight three examples: the writing process graph, the summary analysis and the pause analysis.

The *writing process graph* generates a visual representation of the dynamics of the writing process, including process, product, position, pause, and source information. It allows researchers to quickly orient on how the writing process has been organized and how the text has gradually evolved. Figure 3 shows two examples of such a graph. Both graphs represent a visualization of a writing process in which a synthesis text has been written, based on pre-defined sources.

When we compare both graphs, we immediately notice some differences. First, although the final text length — see lower line — is about 2400 characters (or 470 words) in both cases, the x-axis shows that the second writer needed about double as much time to complete the text: 27 minutes versus 43 minutes. Moreover, the upper line — indicating the amount of characters produced during the writing process — shows that in Case 1 about 85% of the text produced is retained, while in Case 2 less than 40% is retained. This indicates that in the latter case the writer has revised her text much more intensively. The dotted line — indicating the cursor position at each moment in the process — clearly shows that the second writer starts revising her text after about 22 minutes. The dotted line goes down, indicating that she has repositioned the cursor to the beginning of the text and that at that stage she starts revising the text produced so far systematically and in different rounds (five cycles). In other words, the second half of her writing process is almost completely devoted to revising her text. The first writer also starts revising her text after about 23 minutes, but this revision only takes about 4 minutes, or 15% of her total process time.

Finally, the bottom of the graph also shows the interaction with the digital sources that were available. When the line is on top, the writer is consulting sources, when down she is producing text in her Word document. The writer in Case 1 starts writing almost immediately, after briefly having consulted the different sources. From then on, she consistently interacts with the sources, mainly during very short episodes. Only in the final revision period no sources are consulted any more. The second writer shows a completely different interaction pattern. During the first eight minutes she almost exclusively devotes to reading the sources, only writing a few keywords in her text. From then on, she starts producing her own text, regularly checking the sources very briefly. In contrast with the first writer, she keeps checking the source materials during the first part of the revision phase, but not during the final part.

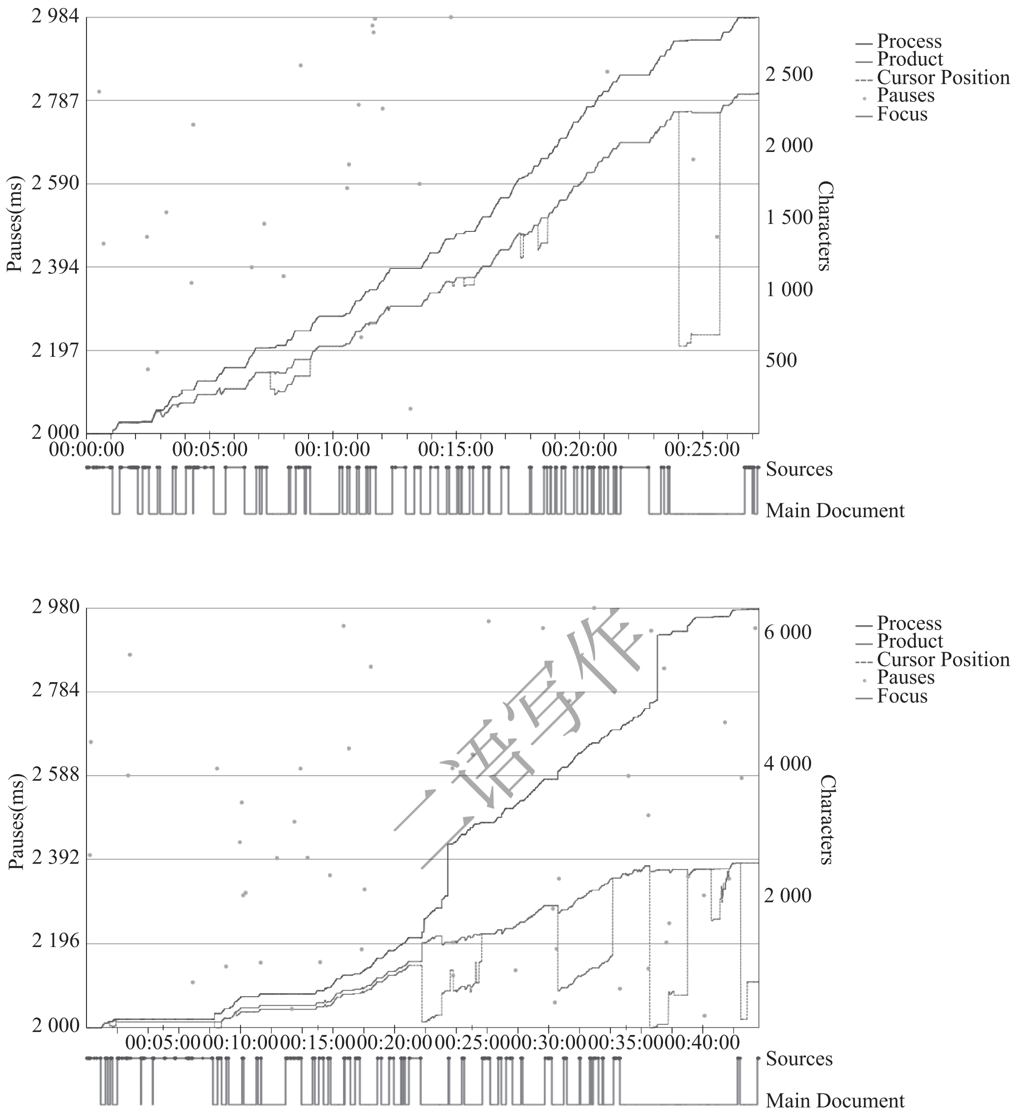

Figure 3. Process graphs for two writing processes (task: synthesis based on sources)

Of course, more details can be spotted in these graphs, but we hope that we have clearly demonstrated that this visualization is a rich and layered starting point to describe or reconstruct the process dynamics characterizing writing processes. While this writing process graph is useful in qualitative case studies (and to support writing pedagogy, see Section 5), other analyses are more suitable for quantitative studies, e.g., comparing participant groups or different writing tasks. For instance, the graph in Figure 4 compares L1 versus L2 writing processes of advanced L2 students for four fluency dimensions (Van Waes & Leijten 2015). Each component is constructed based on a set of underlying variables taken from one of the Inputlog analyses. The figure shows that writing fluency is a multidimensional concept and gives an indication on which dimensions L1 writing processes differ from L2 writing processes — and to what extent. For the first and the last component, the variables are mainly derived from the summary and pause analysis. These two analyses are presented here in short.

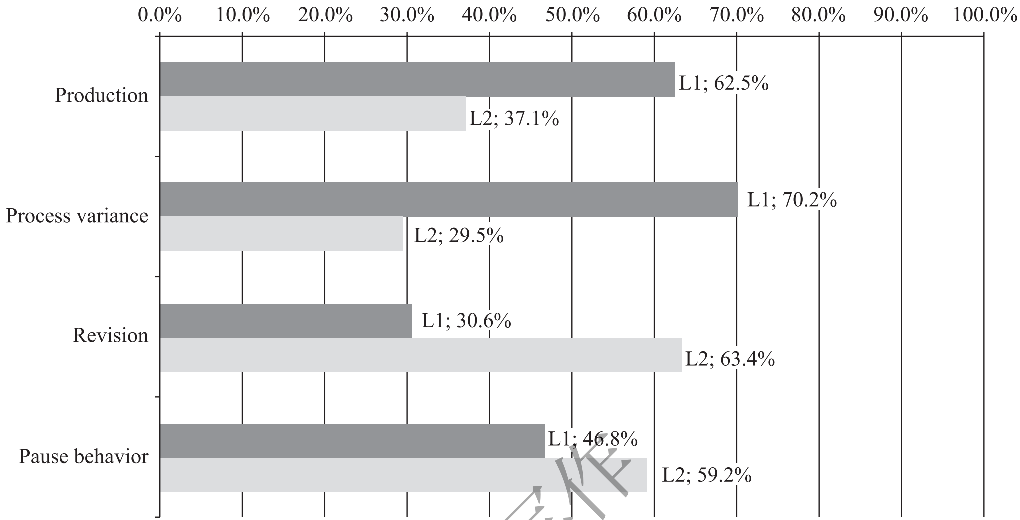

Figure 4. Comparison of L1-L2 fluency dimensions (Van Waes & Leijten 2013)

The *summary analysis* provides a set of aggregated indicators to describe some process and product characteristics of the logged writing process. Table 5 highlights some parts of the output that is generated by this summary analysis. In the example given, the writer produced a final text of 355 words (or 2,277 characters). However, in total she typed 452 words (or 3,104 characters), which means that about 100 words were deleted during the process. The product-process analysis consequently reports a 0.75 ratio (i.e., [Total characters in product + Total non-characters] / Total characters in process = (2,277+50)/3,104 = 0.75), or — when not considering the number of copied characters — a 0.79 proportion. Other measures not included in Table 5 are total process time, total pausing time, number, and length of P-bursts (see, e.g., Chenoweth & Hayes 2001), and a writing modus analysis (see Inputlog manual for a detailed description).

The final analysis we would like to highlight is the *pause analysis*. It zooms in on the writer's pausing behavior and provides a set of indicators to define pausing from different perspectives. Table 6 shows two small extracts from the output. Applying a pause threshold of 200 ms the output reports that 24:51 out of a total process time of 42:44 were spent pausing, which is about 60% of the total time. This pausing time is distributed over 1,154 pauses which have a median duration of 412 ms. As we know that pauses are not normally distributed and are right skewed (see Figure 5), we also report a geometric mean of pauses based on a logarithmic conversion (trimming the 2.5% upper and lower boundaries). We consider the geometric mean as a more refined calculation of the pause duration.

Table 5. Extracts taken from the Inputlog summary analysis output (Inputlog 8.0)

Process Information	
Keystrokes in This Session	
Total Keystrokes in Main Document	3104
- Total Non-Character Keys	50
- Characters Inserted	3
- Characters Replaced	168
- Total Typed (incl.spaces)	2883
- Per Minute (incl. spaces)	67.45
- Total Typed (excl.spaces)	2400
- Per Minute (excl.spaces)	56.15
Words	
Total Words in Main Document	452
Per Minute	10.575
Mean Word Length	5.177
Median Word Length	4
Standard Deviation Word Length	3.425

Product Information	
Characters in Final Text of This Session	
Total (incl.spaces)	2277
Per Minute (incl. spaces)	53.272
Total (excl.spaces)	1912
Per Minute (excl.spaces)	44.733
Words	
Total Words in Main Document	355
Per Minute	8.306

Product/Process	
Ratio	
Produced Ratio (incl.spaces)	0.75
Proportion	
Characters (incl.spaces)	0.79
Words	0.78

Table 6. Extracts taken from the Inputlog pause analysis — Threshold 200 ms (version 8.0)

General Information	
Overview	
Total Process Time	0:42:44
Total Pause Time	0:24:51
Total Pause Time (s)	1491.101
Total Number of Pauses	1154
Arithmetic Mean of Pauses (s)	1.292
Median Pause Time (s)	0.412
Geometric Mean of Pauses (s)	0.584
95% Low Boundary (s)	0.551
95% High Boundary (s)	0.620
Coefficient of Variation	138.06%
Standard Deviation (s)	2.735

Pause Location	
Within Words	
Number of Pauses	256
Arithmetic Mean of Pauses (s)	0.349
Median Pause Time (s)	0.276
Geometric Mean of Pauses (s)	0.311
Standard Deviation (s)	0.261
Before Words	
Number of Pauses	159
Arithmetic Mean of Pauses (s)	0.686
Median Pause Time (s)	0.386
Geometric Mean of Pauses (s)	0.465
Standard Deviation (s)	0.932

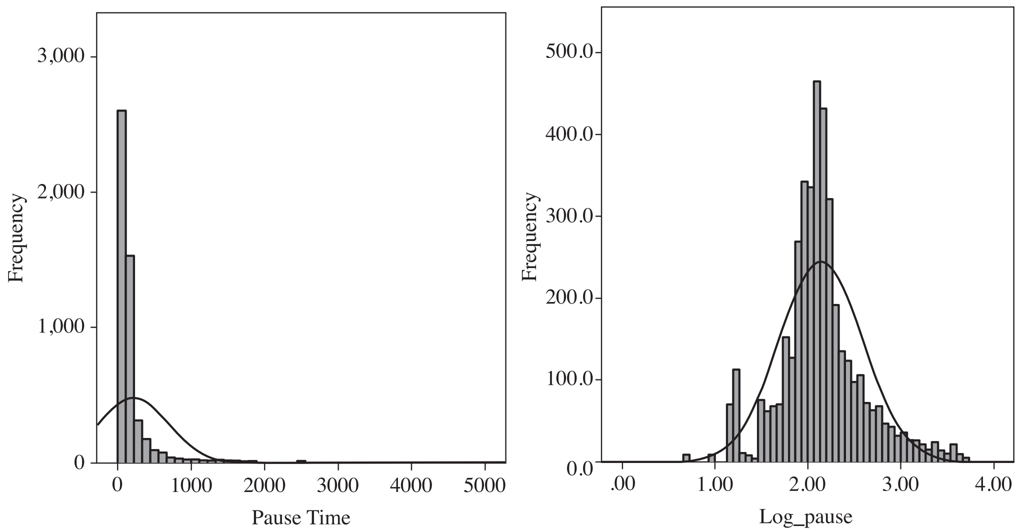

Figure 5. Histogram for pause distribution | left: absolute pause time — right: log converted pause time

In another part of the pause analysis the pauses are classified at different text levels, viz., word, sentence, and paragraph level. The output shows, for instance, that 256 pauses — or 22% of the total pauses — are situated within words. As we mentioned before, the pause analysis is focused on the writer's cognitive load. Therefore, we used a threshold of 200 ms so as to exclude most of the fluent interkey transitions that are mainly motorically defined (Remark: The copy task analysis indicated that this writer had very good typing skills of about 495 characters per minute; mean interkey interval: 109 ms, *SD*=69). For this person it means that about 10% of the total within word interkey intervals are above the 200 ms threshold. From previous research we know that these pauses might be related to, for instance, spelling problems or typo corrections (Conijn et al. 2019). On average the within word pause duration is 311 ms (geometric mean). The pause duration gradually increases at higher text levels (see Figure 6). We also know that the distribution of pause duration and distribution within each category is more complex than this more general approach presented here. For a more detailed perspective we refer to Galbraith and Baaijen (2019) who state that the distribution of pause duration at the between and within word level is characterized by a multimodal nature of the distributions, showing two, or sometimes three, different distributions.

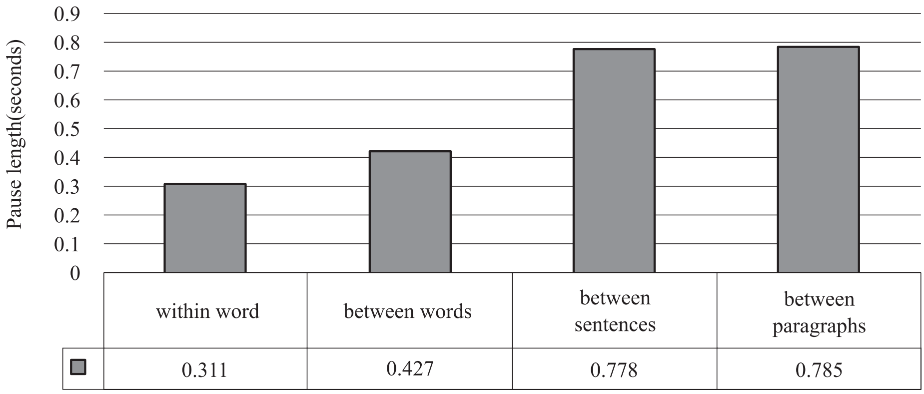

Figure 6. Mean pause duration at different text levels

This brief introduction to three analyses aimed at illustrating that different perspectives to keystroke logging data are needed to fully address the richness of data collection and describe the underlying cognitive process. Bringing together diverging and complementary perspectives makes it easier to fully grasp the complexity of the dynamics of writing processes. However, the examples given also clearly indicate that it is not always easy to infer solid interpretations from the analyses. Therefore, we contend that more research is highly needed, both theoretically and practically oriented.

5. Conclusion and Future Perspectives of Keystroke Logging

As the literature analysis (Section 2) indicates, keystroke logging has become a mainstream research method in writing (and translation) research. Keystroke logging is mostly used in more applied studies, and to a lesser extend for explicit theory building. In recent years, the number of methodological papers increases since the techniques of analyzing keystroke logging data become more advanced. Keystroke logging data can be processed more easily from various perspectives like source use (Leijten et al. 2019), fluency (Van Waes & Leijten 2015) and typing skills (Van Waes et al. 2019). Additionally, the statistical techniques become more advanced and complex such as data mining (Sinharay et al. 2019; Warschauer et al. 2019) and Bayesian linear mixed effects models (Conijn et al. 2019). Moreover, the overview shows the wide variety of research domains in writing that successfully incorporated this type of observation method via Inputlog.

Next, this article describes the research flow for keystroke logging studies in more detail. This description is linked to Inputlog and the different modules this program provides to guide scholars through this research process. We hope that the concrete description of how to design a keystroke logging study will inspire readers as researchers and will help them in setting up well-controlled and rigorous studies. The final section of this article provides some examples of log data analysis illustrating a few perspectives on the exploration of keystroke logging data.

Keystroke logging has only been developed quite recently and current research shows that there are ample possibilities to further develop and strengthen the use of this research method. The close collaboration between the developers of the existing log programs and researchers in this domain is important in this perspective. To end up this article we briefly present two ongoing projects we are involved in as kind of preview of further development.

First, we are developing a so-called “report”-function in Inputlog that targets the use of keystroke logging in pedagogical contexts (Vandermeulen et al. 2020). More specifically, as most feedback now is product oriented, we contend that it might be worthwhile providing pupils and students with adequate feedback about their writing process. Because more and more writing — also in education — is digitized, keystroke logging tools could facilitate this type of feedback. In the prototype we have been developing, we create a report that includes process indicators derived from the existing Inputlog analyses together with a brief explanation and interpretation of the data presented. We include both graphical and numeric information. These reports are automatically generated and should function as an easy-to-interpret feedback to reflect on the way in which a student has organized his or her writing process: which writing episodes were more fluent than others? To what extent was the process linearly developed? How did the student interact with sources (e.g., use of dictionary)? Also, when students are invited to compare their process reports, we think this might deepen their understanding of how to efficiently and effectively organize their writing process. In the current stage of this development we are piloting different report versions, and we hope to report on the findings of intervention studies soon.

Another orientation of our current research deals with the possibility to not only log Western characters, but also Chinese characters. Together with our colleagues at Shandong University (principal investigator: Prof. Junju Wang), we have developed a prototype of an Inputlog module that matches the *Sogou* pinyin input to the resulting Chinese character. We have opted for the Sogou input method since this is most widely used (Wang et al. 2018). This module should allow researchers to compare, for instance, L1 Chinese writing processes to L2 English writing processes. The main challenges in the current stage relate to methodological and technical issues. From a theoretical perspective we need to redefine certain concepts like those that relate to pausing behavior (at different levels), revision, and writing fluency (Lu 2020).

To sum up, we contend that keystroke logging is a powerful research observation method, and we are sure that in the coming years the tools and analyses will be further developed. New theoretical insights will increase our understanding of the underlying cognitive aspects of writing processes. We hope that this introduction will inspire more researchers to systematically build on a common research agenda pushing the domain of writing research further forward.

References

Brandt, D. 2015. *The Rise of Writing: Redefining Mass Literacy* [M]. Cambridge: Cambridge University Press.

Carl, M. 2012. *Translog-II: A Program for Recording User Activity Data for Empirical Reading and*

Writing Research [R]. Paper presented at the LREC.

Chenoweth, N. A. & J. R. Hayes. 2001. Fluency in writing: Generating text in L1 and L2 [J]. *Written Communication* 18(1): 80-98.

Conijn, R., Roeser, J. & M. Van Zaanen. 2019. Understanding the keystroke log: The effect of writing task on keystroke features [J]. *Reading and Writing* 32(9): 2353-2374.

Conijn, R., M. Van Zaanen, M. Leijten & L. Van Waes. 2019. How to typo? Building a process-based model of typographic error revisions [J]. *The Journal of Writing Analytics* 3: 69-95.

Flower, L. & J. R. Hayes. 1981. A cognitive process theory of writing [J]. *College Composition and Communication* 32(4): 365-387.

Galbraith, D. & V. M. Baaijen. 2019. Aligning keystrokes with cognitive processes in writing [A]. In E. Lindgren & K. P. H. Sullivan (eds.). *Observing Writing: Insights from Keystroke Logging and Handwriting* [C]. Leiden/Boston: Brill. 306-325.

Hayes, J. R. 1996. A new framework for understanding cognition and affect in writing [A]. In C. M. Levy & S. Ransdell (eds.). *The Science of Writing: Theories, Methods, Individual Differences, and Applications* [C]. Mahwah: Lawrence Erlbaum Associates. 1-27.

Hayes, J. R. 2012. Modeling and remodeling writing [J]. *Written Communication* 29(3): 369-388.

Hyland, K. 2016. Methods and methodologies in second language writing research [J]. *System* 59: 116-125.

Hyland, K. 2019. *Second Language Writing* [M]. Cambridge: Cambridge University Press.

Johansson, V., Å. Wengelin, J. Frid & R. Johansson. 2014. *Scriptlog 2013: State of the Art* [R]. Paper presented at the training school on keystroke logging, University of Antwerp, Belgium.

Leijten, M., S. Bernolet, I. Schrijver & L. Van Waes. 2019. Mapping master's students' use of external sources in source-based writing in L1 and L2 [J]. *Studies in Second Language Acquisition* 41(3): 555-582.

Leijten, M., E. Van Horenbeeck & L. Van Waes. 2019. Analyzing keystroke logging data from a linguistic perspective [A]. In K. Sullivan & E. Lindgren (eds.). *Observing Writing: Insights from Keystroke Logging and Handwriting* [C]. Amsterdam: Brill. 71-95.

Leijten, M. & L. Van Waes. 2006. Inputlog: New perspectives on the logging of on-line writing processes in a Windows environment [A]. In K. P. H. Sullivan & E. Lindgren (eds.). *Computer Keystroke Logging and Writing: Methods and Applications* [C]. Oxford: Elsevier. 73-94.

Leijten, M. & L. Van Waes. 2013. Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes [J]. *Written Communication* 30(3): 358-392.

Leijten, M. & L. Van Waes. 2019. *Manual Inputlog 8.0* [OL]. https://www.inputlog.net/downloads/ (accessed 28/09/2020).

Leijten, M., L. Van Waes, K. Schriver & J. R. Hayes. 2014. Writing in the workplace: Constructing documents using multiple digital sources [J]. *Journal of Writing Research* 5(3): 285-337.

Lindgren, E., M. Leijten & L. Van Waes. 2011. Adopting to the reader during writing [J]. *Written Language and Literacy* 2: 188-223.

Lindgren, E. & K. P. H. Sullivan (eds.). 2019. *Observing Writing: Insights from Keystroke Logging and Handwriting* [C]. Amsterdam: Brill.

Lu, X. 2020. *Writing in a Non-alphabetic Language Using a Keyboard: Behaviours, Cognitive Activities and Text Quality* [D]. PhD dissertation. London: University College London.

Mackey, A. & S. M. Gass. 2015. *Second Language Research: Methodology and Design* [M]. New York/London: Routledge.

Olive, T. 2014. Toward a parallel and cascading model of the writing system: A review of research on writing processes coordination [J]. *Journal of Writing Research* 6(2): 173-194.

Schrijver, I., L. Van Vaerenbergh, M. Leijten & L. Van Waes. 2014. The translator as a writer: Measuring the effect of writing skills on the translation product [A]. In D. Knorr, C. Heine & J. Engberg (eds.). *Methods in Writing Process Research* [C]. Frankfurt am Main: Peter Lang. 99-122.

Schrijver, I., L. Van Vaerenbergh, M. Leijten & L. Van Waes. 2016. The impact of writing training on transediting in translation, analyzed from a product and process perspective [J]. *Perspectives* 24(2): 218-234.

Sinharay, S., M. Zhang & P. Deane. 2019. Prediction of essay scores from writing process and product features using data mining methods [J]. *Applied Measurement in Education* 32(2): 116-137.

Van Waes, L. & M. Leijten. 2006. Logging writing processes with Inputlog [A]. In L. Van Waes, M. Leijten & C. Neuwirth (eds.). *Writing and Digital Media* [C]. Oxford: Elsevier. 158-165.

Van Waes, L. & M. Leijten. 2015. Fluency in writing: A multidimensional perspective on writing fluency applied to L1 and L2 [J]. *Computers and Composition* 38: 79-95.

Van Waes, L., M. Leijten, T. Pauwaert & E. Van Horenbeeck. 2019. A multilingual copy task: Measuring typing and motor skills in writing with Inputlog [J]. *Journal of Open Research Software* 7: 1-8.

Vandermeulen, N., M. Leijten & L. Van Waes. 2020. Reporting writing process feedback in the classroom: Using Keystroke Logging data to reflect on writing processes [J]. *Journal of Writing Research* 12(1): 109-139.

Wang, Z., Y. Guo, S. Zheng, W. Xu, L. Liu, Z. Liu & X. Cui. 2018. Users' location analysis based on Chinese mobile social media [J]. *Concurrency and Computation: Practice and Experience* 32(13): e4669.

Warschauer, M., S. Yim, H. Lee & B. Zheng. 2019. Recent contributions of data mining to language learning research [J]. *Annual Review of Applied Linguistics* 39: 93-112.

Wengelin, Å., M. Torrance, K. Holmqvist, S. Simpson, D. Galbraith, V. Johansson & R. Johansson. 2009. Combined eyetracking and keystroke-logging methods for studying cognitive processes in text production [J]. *Behavior Research Methods* 41(2): 337-351.

Acknowledgments

The development of the Inputlog 9 beta version was financially supported by the Research Foundation-Flanders (FWO). We would like to thank Junju Wang and Wei Cheng (Shandong University) for collaborating in this project. We are also grateful to Cuiqin Xu (Nanjing University) who helped us in selecting Chinese keystroke logging publications.

About the Authors

Mariëlle LEIJTEN (corresponding author) is an associate professor at the University of Antwerp, Department of Management. Her research focuses on cognitive writing processes. Email: marielle.leijten@uantwerpen.be.

Luuk VAN WAES is a full professor at the University of Antwerp, Department of Management. He has been involved in different types of writing research, with a focus on digital media and (professional) writing processes.

征稿启事

《二语写作》是国内第一本关于二语写作教学与研究的学术集刊。由中国英汉语比较研究会写作教学与研究专业委员会主办，外语教学与研究出版社协助出版，编辑部设在山东大学外国语学院。

一、办刊宗旨

《二语写作》重点探讨二语写作理论、写作修辞、写作教学、写作教师教育、写作测试、ESP/EAP写作、机辅写作、技术与写作等相关领域的研究问题。本刊接受中、英文稿件，欢迎不同语境下的研究成果，除英语外，也刊登非英语和对外汉语等领域的二语写作研究成果。本刊倡导理论创新，关注研究方法，强调学术的高度、深度和厚度，尤其注重跨学科视角下的研究与探讨，旨在为我国二语写作研究学者、教师和研究生提供学术交流的平台，推动我国二语写作教学与研究的发展。

二、重点栏目

本刊设有“理论视角”、“教学研究”、“测试与评估”、“跨学科研究”、“新秀论坛”、“学术动态”、“新作评介”、“注释文献”等重点栏目，真诚欢迎国内外专家、学者和广大外语教师及研究者赐稿。

三、稿件要求与投稿说明

1. 来稿应具有科学性、原创性，论点鲜明、论据充分、数据准确、逻辑严谨、文字通顺、图表规范。每篇论文以8000字左右为宜，最多不超过1万字，中英文均可。

2. 来稿请详细注明作者简介、作者单位、地址、邮编、联系电话及电子邮箱。稿件不涉及保密问题，署名无争议，稿件一律文责自负，本刊有权对来稿作文字修改。

3. 本刊恕不退稿，请作者自留底稿，切勿一稿多投。

4. 请不要把稿件邮寄给个人，以免影响审稿和发表。

5. 按照国际惯例，《二语写作》实行同行专家匿名审稿制度。审稿周期为3个月。3个月后，若没有收到编辑部反馈，作者可自行处理。

6.《二语写作》不收版面费。

四、投稿方式

1. 电子投稿：请直接发送至writing-journal@sdu.edu.cn

2. QQ群：232248241、976907510、222462129

二语写作

Chinese Journal of Second Language Writing

中国英汉语比较研究会写作教学与研究专业委员会主办
山东大学二语写作教学与研究中心承办

《二语写作》第一辑

（2020 年 10 月）

目　录