

# Writing Fluency and Digital Source Use in Authentic Academic Writing

Sergej Ivanov, Annelie Johansson & Christian Waldmann

## Abstract

This paper presents a case study that uses keystroke logging to explore writing fluency and digital source use during an authentic academic writing process in a novice student of data science through the lens of the Simple View of Writing. The study found that the writing of the literature review lasted for 11 hours distributed over 12 sessions. The sessions were thematically divided into four well-defined units with only minor overlaps that could be explained by the design of the task and that may reflect the non-linear and recursive implementation of writing processes at the text production, transcription and monitoring level. The findings suggested that the writer's fluency, as measured by pauses, revisions and production rate, as well as digital source use reflect the cognitive effort involved in academic writing. The combined pause and revision data indicated a flexible writing profile depending on the complexity of the task at hand and on any approaching deadline for submission of a draft and receipt of feedback. The student demonstrated an effective source use in writing both for informative purposes in the results section and for argumentative purposes in the introduction and discussion. The findings may inform the teaching and learning of academic writing in terms of time allocation for the associated tasks and pedagogical support that addresses crucial language and genre knowledge as well as facilitates the text production and efficient source use.

## Introduction

Learning how to write academic texts is vital for success at university studies, and academic writing skills are requested in many professions outside the universities. Writing in general is a complex and cognitively demanding task that involves low-level cognitive processes supporting transcription (e.g. spelling, handwriting, keyboard control) as well as high-level processes supporting text production (e.g. knowledge about the topic and language) and monitoring the writing process (e.g. planning, evaluating and revising the written text in relation to task requirements) (Berninger & Amtmann 2003). Academic writing further adds to this cognitive challenge with its specific language, genre, structural and formatting requirements, and procedures for synthesizing large quantities of text sources and critically evaluating different viewpoints. At Swedish universities, there is an increasing demand for supporting students' development of academic

writing skills. The effectiveness of such support depends on an adequate and detailed knowledge about the process of academic writing, which qualifies as a research problem in educational linguistics.

In writing research, keystroke logging is a method that uses a computer software to record keystrokes, mouse actions and the use of sources during writing logged on a timeline, which has the potential to result in valuable knowledge about the writing process (e.g. Lindgren & Sullivan 2019). This method builds on insights from cognitive writing research that writing behaviours reflect cognitive processes during writing (e.g. Berninger & Amtmann 2003; Chenoweth & Hayes 2001; Flower & Hayes 1981; Hayes 2012). The writing behaviour that is the focus of this study is writing fluency, which has been described as “the end product of all the writing processes” (Olive, Favart, Beauvais & Beauvais 2009: 305). Writing fluency is multifaceted and shaped by the writer’s pausing and revision behaviour and production rate (MacArthur, Graham & Fitzgerald 2008; Van Waes & Leijten 2015).

Previous studies using keystroke logging to examine academic writing have typically used short writing tasks in experimental designs to study one specific aspect of the writing process, and few studies have examined a longer, authentic academic writing process (see Bowen & Van Waes 2020). To our knowledge, there are no previous studies taking a holistic approach to the academic writing process using keystroke logging to explore several aspects of an authentic academic writing process as it progresses over numerous sessions. The aim of the present case study is to explore the writing fluency and digital source use in an authentic academic writing process consisting of several writing sessions in a novice student of data science. The following research questions are addressed:

- What is the overall thematic structure of the writing sessions?
- What characterises pausing, revision and production rate during the sessions?
- What characterises the use of digital sources during the writing sessions?

The study contributes new and valuable knowledge about the academic writing process that can inform the teaching and learning of academic writing at university level.

The article is structured as follows: The next section outlines the theoretical background and presents previous research on writing fluency and source-based writing. The following sections present the methodology and the results of the study. Finally, the results are discussed and implications for education are presented.

## Background

According to an influential model of the writing process, the Simple View of Writing (SVW), writing involves cognitive processes at three levels – text production, transcription and monitoring – coordinated by the writer’s working memory (Berninger & Amtmann 2003). Working memory supports text production, transcription and monitoring by activating and retrieving knowledge from long-term memory and making it temporarily accessible and usable for the writer (Kellogg, Whiteford, Turner, Cahill & Mertens 2013). At the text production level, the writer is using their knowledge about, for example, the topic, words, grammatical structures and the specific genre to generate ideas and convert them into words, sentences and a coherent discourse in the head. At the level of transcription, the writer is using knowledge about the phonological, orthographic and morphological structure of words to convert the language form in the head into a written text by spelling words correctly. The writer is also using motor knowledge and skills to write by hand or on a keyboard. To monitor the writing process, the writer is using knowledge about and strategies for planning, evaluating and revising the text in relation to the context of the writing task and the target reader. Processes at the three levels are implemented non-linearly and recursively during writing. The writer responds to aspects such as the text produced so far, the communicative goal, text sources and so on, and continually moves between text production, transcription and monitoring.

Although implicit in the SVW, reading plays an important role in writing. According to Hayes (1996), the quality of the text is often dependent on the writer’s ability to read to evaluate the text produced so far, to read to define the writing task and to read text sources. Reading to evaluate is a prerequisite for revising the text. In other words, writers’ ability to effectively revise the text produced so far depends on their ability to read and understand their own text. Reading to define the writing task is vital for the monitoring of the writing process. To solve an academic writing task, writers need knowledge about the specific language, genre, structure and formatting requirements as well as various skills, such as being able to summarise, synthesise, argue, evaluate, categorise and reflect. Writers’ ability to adhere to such requirements may, at least to some extent, depend on their ability to read and understand various texts that assist them in defining the writing task, such as course materials, example texts and handbooks. Finally, reading text sources plays an important role in solving many types of writing tasks, especially an academic writing task. Text sources can support writers in the process of generating ideas by providing them with information and knowledge about the topic. Writers’ ability to correctly represent information from text sources in their own text depends on their ability to read, understand and interpret the text sources.

The SVW implies that the writer’s cognitive processes are reflected in writing behaviours (e.g. Berninger & Amtmann 2003; Chenoweth & Hayes

2001; Flower & Hayes 1981; Hayes 2012). Academic writing, and the writing of academic reports specifically, is a cognitively demanding task that does not only require knowledge about the specific language, genre and formatting characteristics, but also involves, for example, searching for information about the topic, reading about the topic, synthesising a large number of text sources, critically evaluating various viewpoints and a step-by-step organisation of the content (Matsuhashi 1981). The cognitive demands of academic writing are, thus, likely to be reflected in the writing fluency as measured by pausing, revision and production rate as well as in the way the writer is using various sources. The following sections summarise some previous research on writing fluency and the use of sources.

### **Writing fluency**

Fluent writing refers to a state of effortless writing characterised by “short pausing times, few revisions and a high production rate” (Van Waes & Leijten 2015: 80). Thus, the writer’s pausing and revision behaviours as well as the rate with which the writer is producing the text contribute to shaping fluency in writing. Frequent pausing and revision, long pauses, lengthy revisions and a slow production of the text itself reduce the writing fluency. This section introduces pausing, revision and production rate as key features of the concept of writing fluency.

### ***Pausing***

Pauses are temporal breaks in the writing activity, that is a scriptural inactivity, that arise, on a computer, when no keys are being pressed. A writer may pause for various reasons that may or may not be associated with the writing task. Some pauses arise due to motor activities involved in writing, such as the time it takes to move the fingers to the next key or combination of keys and move the hand between the mouse and the keyboard (Olive & Kellogg 2002). Studies examining the impact of writer’s transcription skills on writing fluency have found that fluency is enhanced by automatised spelling and motor skills for writing by hand or on a computer, for example, finding and pressing the correct keys (e.g. Alves & Limpo 2015; Connelly, Dockrell, Walter & Critten 2012; Sumner, Connelly & Barnett 2013; Wengelin, Johansson & Johansson 2014).

Writing research using keystroke logging is mainly concerned with pauses arising as a result of cognitive processes during writing, such as planning, formulating, reading and evaluating (e.g. Alves & Limpo 2015; Flower & Hayes 1981). There is no one-to-one relation between a particular writing behaviour, such as a pause, and a specific cognitive process. A pause may reflect different cognitive processes, and a particular cognitive process may be reflected in different writing behaviours (Schilperoord 2002).

One challenge is to identify pauses that are likely to reflect cognitive processes. A factor that has proven particularly promising for distinguishing

between different pauses is the duration of the pause. Whereas pauses that arise due to motor activities are brief (usually only a few hundred milliseconds), pauses that arise when the writer engages in cognitively more demanding activities are longer (Olive & Kellogg 2002; Schilperoord 1996). A common threshold for distinguishing between these types of pauses among adults, that is also applied in this study, is 2,000 milliseconds (e.g. Chenu, Pellegrino, Jisa & Fayol 2014). In other words, pauses longer than 2,000 milliseconds are assumed to exclude pauses due to motor activities.

A complicating factor is that writers may also pause when involved in activities not directly related to the writing task, such as thinking of a soccer game, looking out the window, or drinking a glass of water. This is particularly so if the writing task stretches over several days and many long sessions, as in the current study. Leijten, Van Waes, Schriver and Hayes (2014) argued, however, that such downtime plays an important role when solving longer and more complicated writing tasks. Such pauses may reflect a writer's "meta-knowledge of their own motivational limits" (Leijten et al. 2014: 331) and may be important to relieve fatigue and increase concentration on the task at hand.

Research on pausing behaviour among writers has focused on, among other things, pausing at different syntactic locations (within and between words, sentences, clauses and paragraphs), as well as before and after punctuation and spaces (e.g. Matsushashi 1981; Spelman Miller 2000; Wengelin 2006). Pauses between segments of the text may arise as a result of planning how to continue the text that may or may not be based on reading and evaluating the text produced so far and reading text sources (Baaijen, Galbraith & de Glopper 2012). One finding is that planning at a macro-level (e.g. between sentences and paragraphs) results in longer pauses than planning at a micro-level (e.g. between words) (e.g. Spelman Miller 2000). Pauses arise during writing when the writer activates conceptual and linguistic knowledge to accomplish writing goals (e.g. Schilperoord 1996; Spelman Miller 2006). Such activation of knowledge is to different degrees constrained by the previously activated knowledge. For example, activation that concerns micro-level planning is more constrained by the previously activated knowledge, and the shorter pauses associated with micro-level planning reflect the need for less cognitive effort to activate a piece of knowledge for the progression of the writing. In contrast, planning at a macro-level (e.g. transitions between topics) is constrained by the previously activated knowledge to a lesser degree and requires more cognitive effort and maybe also reading and evaluating (parts of) the text produced so far and reading other text sources, and hence longer pauses, to activate the conceptual and linguistic knowledge required for the progression of the writing.

As regards academic writing, the specific way of writing a report by, for example, a step-by-step organisation of the content, searching for, reading

and synthesising a large number of text sources and critical evaluation demands cognitive effort and, thus, time for advanced planning, reading and evaluating at the macro-level (Matsuhashi 1981). This is likely to be reflected in pausing behaviours, and as a consequence in writing fluency, when writing an academic report.

### **Revisions**

Revisions are changes to the text – such as additions, deletions and substitutions – that play an important role in the process of writing. A writer may engage in different types of revisions that can occur either internally in the head or externally in the text (Lindgren & Sullivan 2006). Internal revisions may happen prior to converting ideas into language in the head and concern conceptual revisions of plans and ideas, or prior to transcription and concern revisions of concept and form. Although internal revisions occur mentally and are not directly visible in the text, they may be represented by pauses during writing. External revisions, however, are directly visible in the text as changes to the concepts (e.g. content, meaning), form (e.g. spelling, grammar) or typography of the written text. Such revisions may or may not consider the context of the text produced so far. A writer may revise independently of the context in parallel with the transcription of the text at the point of inscription (Chenoweth & Hayes 2001). Alternatively, a writer may stop the transcription process to read, evaluate and revise the text produced so far. This study mainly concerns external revisions that become directly visible in the keystroke logging data. However, the pausing behaviour of the writer may partly reflect internal revisions.

Revision behaviours, and as a consequence writing fluency, vary as a function of language proficiency and writing experience (Chenoweth & Hayes 2001; Kobayashi & Rinnert 2013; Lindgren, Leijten & Van Waes 2011; Lindgren, Sullivan & Spelman Miller 2008; Van Waes & Leijten 2015; see also Lindgren & Sullivan 2006). L2 writers revise more often than L1 writers, and like inexperienced writers, they focus more on the linguistic form when revising. An efficient and fluent retrieval and use of adequate language knowledge reduce the need for revisions and enhance the writing fluency.

Revision behaviours may also vary as a function of the personality of the writer. Studies on revision patterns have identified two main writing profiles that Galbraith (1999) calls high-self monitors and low-self monitors. High-self monitors devote much time to advanced planning and idea generation prior to transcribing. Consequently, they take longer pauses in the initial phase of writing and make fewer revisions (e.g. Van Waes & Schellens 2003). Low-self monitors, on the other hand, spend less time on initial planning and idea generation. They approach writing as “sculptors” and use writing as a tool to create the content of the text. Thus, low-self monitors

make many short pauses and revisions during writing, and revisions are particularly frequent when writing the first draft.

Research on revisions during the writing of longer texts that are developed through several drafts written over several sessions in authentic (non-experimental) settings is scarce. Bowen and van Waes (2020) present a pioneer study of revisions in one British undergraduate student (19 years old) during the writing of 3 essays that constitute the assessments within an English language program. The essays were written over 20 sessions (almost 7 hours). The writing process was recorded using a keystroke logging software (Inputlog), and the writing processes and text products were analysed to examine the functions of revisions as well as the time and place of revisions. They found that the student had the characteristics of a low-self monitor who creates the content of the text during the writing. The main part of the content was produced in the first two sessions, which were much longer than the subsequent sessions. Revisions were mainly made at, or ahead of, the point of inscription. Text production dropped in the subsequent sessions that were mostly devoted to proofreading, minor additions and refinements of the existing text.

### ***Production rate***

Although writing fluency is generally defined by the speed of execution, it does not simply mean 'writing quickly'. Rather, fluency indicates cognitive constraints during writing and, as evidenced above, the distribution of pausing and revision behaviours during writing contributes to shaping fluency in writing. In addition to pausing and revision, the rate with which the text itself is produced also contributes to forming fluency, and various measures have been suggested to capture different aspects of production rate (Chenoweth & Hayes 2001; Johansson, Wengelin, Johansson & Holmqvist 2010; Van Waes & Leijten 2015). Two such aspects are product- and process-based production rate. Product-based measures consider production rate in relation to the length of the final text and includes measures such as the number of characters or words in the final text. Such product-based measures do not consider that writers may revise and delete characters and words during writing and that such revision behaviours may influence the production rate in various ways. Keystroke logging software allows access to process-based measures of production rate, which considers the revision behaviours of writers and includes measures of the speed of writing, such as the number of characters or words (including revised characters and words) per minute. The current study considers both product- and process-based production rate.

### **Sources-based writing**

The ability to write effectively from sources is key to academic literacy (Cumming, Lai & Cho 2016). Source use includes searching for, reading, understanding and synthesising a large number of text sources as well as

using paraphrasing and citation practices to integrate relevant content into one's own text (Davis 2013; Liu, Lin, Kou & Wang 2016). Source use interacts with pausing, revision and production rate. When using sources during the writing of an academic text, a writer frequently interrupts the text production and continuously alternates between reading various text sources and (re)writing their own text (Mateos, Solé, Martín, Miras & Castells 2014). A writer must simultaneously understand the content of the text sources, choose relevant and a suitable amount of information, present the information logically and coherently in writing and adapt the presentation to the intended audience. Thus, academic writing is a cognitively demanding reading-to-write task that builds on the writer's linguistic skills as well as their reasoning and problem-solving skills (Leijten, Van Waes, Schrijver, Berolet & Vangehuchten 2019; Plakans 2008).

Previous research has focused, among other things, on how language and writing proficiency relate to the use of sources. Students with a lower language proficiency may find it difficult to understand and restructure main ideas in source texts, and they may compensate for these difficulties by copying strings from the source material into their own text (McDonough, Crawford & De Vleeschauwer 2014). Some studies have also found that L2-learners tend to use fewer citations to refer to text sources than L1-learners (Keck 2006; Shi 2004), but this finding is not unequivocal as all students seem to go through similar stages in developing strategies for source-based writing (Hyland 2009; Keck 2014).

Strategies for writing from sources develop as writing experience increases (Cumming, Lai & Cho 2016; Davis 2013; Keck 2014). Early steps in source-based writing includes presenting ideas linearly using verbatim copying from source texts, over-citation and direct quotations (Cumming et al. 2016; Davis 2013). With increasing experience students develop more elaborate strategies that allows knowledge transformation and conceptual integration (Cumming et al. 2016; Keck 2014). Cumming, Kantor, Baba, Erdosy, Keanre and James (2005) found that the most proficient writers tended to summarise and synthesise ideas coherently, whereas middle-range writers used more paraphrasing and plagiarizing. The least proficient writers summarised, paraphrased and copied less than other writers.

Research on the use of sources in academic writing from a process perspective is scarce. The few studies that have been conducted range from secondary school to university. A general finding is that a recursive approach to source use is beneficial for text quality, whereas a more sequential approach is related to lower text quality (e.g. Solé, Miras, Castells, Espino & Minguela 2013). The use of sources may, however, influence text quality differently depending on how and at what point in the writing process sources are used. For example, university students who spend more time reading sources and frequently switch between sources in the prewriting phase are less focused on content elaboration, grammar and

spelling (Escorcia, Passerault, Ros & Pylouster 2017). By contrast, long and attentive reading before writing and frequent switching between sources during writing have been shown to impact positively on text quality (Leijten et al. 2019).

Vandermeulen, van den Broek, Van Steendam and Rijlaarsdam (2020) examined patterns of (in)effective source use in argumentative and informative synthesis writing in upper secondary school and identified differences relating to the temporal distribution and genre that may be relevant when considering source use in academic writing at university level. They found that an effective source use in argumentative writing was characterised by a considerable amount of time spent in sources and frequent transitions between sources and the text at the beginning of the writing process. Little time spent in sources at the end of the writing process also characterised an effective source use. This indicated that proficient writers initially focused on reading sources with the goal to support a position and on selecting information from sources and writing it down. At the end of the writing process, proficient writers focused on writing and revising the text rather than selecting information from sources. Interestingly, students who displayed an ineffective source use and spent very little time or an excessive amount of time in sources in the beginning of the writing wrote texts with lower quality. This highlights the importance of allowing enough time to read and understand sources before starting to write the text. The students who spent an excessive amount of time reading sources likely experienced problems with understanding the sources due to, for example, low reading ability (see also Plakans 2009).

An effective source use in informative writing was characterised by frequent transitions between different sources in the beginning of the writing process (Vandermeulen et al. 2020). To write a high-quality informative text, writers initially need to read, compare and contrast information from different sources in order to identify the overarching theme of the text to be written. At the end of the writing process, proficient writers switch between sources less often. Students who are frequently switching between sources and spending much time in sources at the end of the writing process likely experience problems with integrating information from sources in the text and have lost focus on writing and revising the text.

## **Method**

This section presents the participant and material for the study as well as the methods that were used to collect and analyse the data.

### **Participant**

To recruit a participant, the study was presented orally to a group of approximately 100 students majoring in Computer Science and taking a 7.5-ECTS-credit course in academic writing. The students were invited to

submit their contact details to receive further information. As a result, 5 students expressed their interest in the study and an information sheet and a consent form (see Appendix 1) were distributed to them along with an invitation to a 15-minute tutorial on how to install and use the keystroke logging software. The students also received a research article describing the software (Leijten & Van Waes 2013) and its user manual for optional reading. After the tutorial, two students returned signed consent forms and installed the software on their laptops.

In this case study, the data from one of these students is examined, a 20-year-old male. He will be henceforth referred to as Baldur, a pseudonym that was picked by himself. The other student stopped logging the writing process halfway because of some hardware issues. Baldur can be described as highly skilled in typing on a keyboard and with a self-reported L1 proficiency of English, which is the language used in the writing task. Thus, neither transcription nor language skills should severely constrain the writing of the text.

### **Material**

As a part of the course in academic writing, students are to compose a report of maximum 10 pages in total with a literature review on a self-chosen topic within the field of Computer Science. Typically, students start writing their reports in a mainstream word processor (e.g. Microsoft Word) and by the end of the course transfer it to LaTeX as they learn how to use the latter. At the beginning of the course, students are provided with a Word template consisting of 225 words (1,464 characters including spaces). The template contains the following with short instructions: a title page, a page for abstract and keywords, a page with automatically generated contents, a page for the running text with headings of different levels and other formats as well as pages for references and appendices.

The students' writing process is guided by lectures, workshops and labs as students gradually develop their reports and receive feedback on each of three drafts from the course tutors and peers. They also receive course materials and an example report to assist them in their writing. Before submitting the report for grading, the students participate in a final seminar and receive additional feedback that supports the composition of the final version of the report. The report that is examined in this study was typical in terms of the length (10 pages in total with 6 pages of running text). Its final draft contained 3,382 words and 21,119 characters including blank spaces. In terms of quality, Baldur's report received the second highest grade B (A–F-scale) from a tutor who was not in any capacity involved in the current study and knew nothing about Baldur's participation. It should be noted that the grade was only set on the final draft.

### **Data collection**

Inputlog 8 ([www.inputlog.net](http://www.inputlog.net) – Leijten & Van Waes 2013) was used to record all keyboard and mouse actions performed during the writing of the report. Besides collecting keyboard and mouse actions performed in MS Word stamped on a timeline, Inputlog also records keystrokes in other software as well as focus events during writing, for example, the use of web pages, documents and programmes. This allows analysis of the various digital sources that Baldur is using during writing. Given that the data was collected in an authentic situation, Baldur had full control of Inputlog and could choose when to start and stop logging. After submitting the report, Baldur shared the keystroke logging data with the researchers.

### **Data analysis**

Inputlog offers an analytical module and modules for pre- and postprocessing. In order to prepare the log files for analysis, noise such as pause-time before the first and after the last keystroke was removed using the preprocessing module. The preprocessed log files were submitted for summary, pause, revision, fluency and source analysis provided by the analytical module, and then postprocessed to create excel-files with quantitative data for a wide range of measures.

To examine Baldur's writing fluency, we included several measures of pausing, revision and production rate that were generated by the pause, revision, fluency and summary analyses. The measures are presented in Table 1 below.

Table 1. *Measures of writing fluency*

Key features of writing fluency	Measure	Description	
Pausing	Time on task	Total logged time	
	Pause time	Total time used for pausing longer than 2,000ms	
	Active process time	Total time used to actively work with the task	
	Number of pauses	Number of pauses longer than 2,000ms	
	Pause length	Average pause length for pauses longer than 2,000ms	
Revision	Number of revision events	Number of deletions and insertions	
	Number of revised characters	Number of deleted and inserted characters	
Production rate	Process	Writing rate: productive keystrokes	Number of keystrokes adding or removing characters
		Writing rate: productive keystrokes per minute	Number of keystrokes adding or removing characters per minute
		Typing rate: keystrokes	Number of keystrokes
		Typing rate: keystrokes per minute	Number of keystrokes per minute
	Product	Words in initial and final text	Number of words in the text at the start and end of each session
		Characters in initial and final text	Number of characters in the text at the start and end of each session

As regards pausing, the pause threshold was set at 2,000ms which means that pauses are assumed to mainly reflect high-level writing processes, such as planning, idea generation, reading and evaluating the written text. To some extent, Baldur is pausing for reasons unrelated to the writing task. However, as we have manually analysed all log-files, we can conclude that there are very few pauses of inactivity in the recordings of this type. There may be pauses arising when Baldur is involved in activities not related to the writing task. For clarity, we report writing-unrelated activities that Baldur is involved in during writing in the results section.

Measures of pausing included time on task, pause time, active process time, number of pauses and pause length. The time on task amounts to the total logged time and corresponds to the sum of pause time and active process time. Pause time was calculated by summarizing the time spent on pauses longer than 2,000ms, whereas active process time was calculated by summarizing the time used to actively work with the writing of the text, including pauses reflecting low-level writing processes (i.e. pauses not longer than 2,000ms). The number of pauses was calculated by summarizing the number of pauses longer than 2,000ms, and average pause length by dividing pause time with number of pauses.

The analysis of revision provided data on the writer's revision behaviour, a vital part of monitoring the writing process. Measures of revision included the number of revision events and the number of revised characters. A

revision event may comprise one or several characters deleted or inserted at the same time. The number of revision events was calculated by summarizing the number of performed deletions and insertions in the text prior to the point of inscription. The number of revised characters was calculated by summarising the number of deleted characters and the number of characters inserted in the text.

The analysis of production rate included both process- and product-based measures. Process-based measures focused on writing rate and typing rate. Writing rate provided information about how fast Baldur was pressing the keys on the keyboard when producing the actual text and included the total number of productive keystrokes (i.e. keystrokes that added or removed characters in the text) and the number of productive keystrokes per minute. Typing rate, on the other hand, provided information about how fast Baldur was pressing the keys on the keyboard and included the pressing of keys that did not add or remove characters in the text, such as shift, control, alt and arrows. Typing rate included the total number of keystrokes and the total number of keystrokes per minute. The product-based measures included the number of words and characters in the text at the start and end of each session.

To examine Baldur's digital source use, the source analysis was used to generate data on the various sources that Baldur used for planning and developing the written text. The sources were functionally categorised as the text, formatting, factual, language, course materials, private and other sources. The text comprised the report under construction in Word and LaTeX. Formatting were the formatting functions used in Word, such as inserting and formatting the table of contents and word count. Factual sources were Internet sources that Baldur searched for and used to produce the content of the text, mainly journal articles and books but also news articles and information on the web pages of scientific associations and universities. Language sources were Internet sources that contributed to the language form of the text, for example, thesauruses, dictionaries and lexicon. Language sources also included Internet sources used to find information about, for example, the structure of an academic report and the various functions in Word. Course materials were different resources supporting the writing of the report, such as the task instructions, an example report and a Google document. Baldur also consulted various private and unrelated sources during writing, such as web pages with music and videos, webmail and wikis. Other sources were marginal sources not directly involved in the task, such as error messages, log-in-pages and network diagnostics. In addition to categorizing the sources used, we manually calculated the time spent on each source category.

## Results

This section presents how Baldur wrote his report over 12 sessions focusing on the overall thematic structure of the writing sessions, pausing and revision behaviours, production rate and the use of digital sources. Table 2 presents an overview of the writing sessions including information about the date and duration of each session. To provide a picture of the text produced in each session, the table also presents data on product-based production rate. For overview purposes, the sessions are combined into larger thematic units that reflect the contents of the report: the introductory, result, discussion and overall review units. As with any non-linear writing, there are some overlaps between the units, for example, the end of session 5 is devoted to preparing the first draft of the discussion, while the main part of the session is focused on the results. For clarity, the intermediate submissions and the final seminar are presented in Table 2.

*Table 2. Overview of the writing sessions*

Unit	Session	Date (dd.mm.year)	Duration			Product-based production rate (words/characters)			
			Time on task (h:min:sec)	Per session (%)	Cumulative (%)	Initial text length	Final text length	Difference	
Introductory	1	13.11.19	00:00:11	0.03	0.03	225/1,464	225/1,464	0/0	
	2	14.11.19	00:55:09	8.3	8.33	225/1,464	455/3,040	230/1,576	
	3	14.11.19	00:03:15	0.5	8.83	455/3,040	463/3,105	8/65	
	15 <sup>th</sup> Nov: Submission 1								
Result	4	01.12.19	05:06:25	46.1	54.93	463/3,105	2,103/15,988	1,640/12,883	
	2 <sup>nd</sup> Dec: Submission 2								
	4 <sup>th</sup> Dec: Peer review on Submission 2								
Discussion	5	09.12.19	02:15:59	20.4	75.33	2,103/15,988	3,233/23,404	1,130/7,416	
	6	10.12.19	00:48:18	7.3	82.63	3,233/23,404	3,422/24,638	189/1,234	
	7	11.12.19	00:05:40	0.9	83.53	3,422/24,638	3,372/24,322	-50/-316	
	8	11.12.19	00:07:55	1.2	84.73	3,422/24,638	3,376/24,376	-46/-262	
	13 <sup>th</sup> Dec: Submission 3								
Overall review	9	17.12.19	00:01:18	0.2	84.93	3,376/24,376	3,366/24,310	-10/-66	
	13 <sup>th</sup> January: Final seminar								
	10	14.01.20	00:17:43	2.7	87.63	3,366/24,310	3,422/24,689	56/379	
	11	15.01.20	01:13:58	11.1	98.73	3,422/24,689	3,422/24,759	0/70	
	12	15.01.20	00:09:22	1.4	100	3,422/24,759	3,466/25,813	44/1,054	

In the following section, the sessions and units are described in detail supported by the quantitative data on pauses, revisions and process-based production rate in Table 3 and 4. Before reporting the results in more detail, a note on the data on revisions in sessions 7–9 in Table 3 is needed. For no apparent reason, it was not possible to extract full data sets on revision events and revised characters from the log-files for these sessions. In Table 3, we report the number of revision events and revised characters that we were able to extract from the log-files. The numbers are marked with question marks to indicate the unreliability of the data and to remind the reader that the actual numbers are higher than the ones reported.

**Table 3. Pauses and revisions**

Unit	Session	Pausing			Revisions		
		Active time (hh:min:sec)	Pause time (hh:min:sec)	Number of pauses	Average pause length (sec)	Number of revision events	Revised characters
Introductory	1	00:00:11	00:00:00	0	0.00	0	0
	2	00:32:29	00:22:40	214	6.36	227	2,533
	3	00:02:13	00:01:02	10	6.21	44	1,003
Result	4	03:00:17	02:06:07	948	7.98	657	9,395
	5	01:20:26	00:55:32	516	6.46	371	1,538
Discussion	6	00:33:18	00:14:59	159	5.66	172	928
	7	00:03:36	00:02:03	26	4.77	24 (?)	75 (?)
	8	00:04:15	00:03:39	33	6.66	31 (?)	2 (?)
Overall review	9	00:01:14	00:00:04	1	4.39	3 (?)	0 (?)
	10	00:13:29	00:04:14	46	5.54	52	4,684
	11	00:45:16	00:28:42	244	7.06	95	369
	12	00:07:35	00:01:47	24	4.46	114	33,683
<b>Total</b>		<b>06:44:19</b>	<b>04:20:49</b>	<b>2,221</b>	<b>7.05</b>		

**Table 4. Process-based production rate**

Unit	Session	Writing rate		Typing rate	
		Keystrokes	Keystrokes per minute	Keystrokes	Keystrokes per minute
Introductory	1	-	-	-	-
	2	3,028	54.9	6,364	115.4
	3	147	45.2	513	157.6
Result	4	7,235	23.6	27,990	91.3
	5	8,014	58.9	16,066	118.1
Discussion	6	2,811	58.2	6,806	140.9
	7	286	50.5	635	112.1
	8	378	47.7	818	103.2
Overall review	9	-	-	-	-
	10	76	4.3	1,847	104.3
	11	1,376	18.6	5,190	70.2
	12	342	36.5	1,684	179.5

The introductory unit (sessions 1–3) approximates to 1 hour of the total 11 hours of duration for all sessions. It is likely that a substantial part of time for macro-planning of the report as well as for reading various sources is not reflected in this value. It is specifically evident in the way how Baldur dealt with factual sources that were used to write the introduction by taking a very quick look of several seconds at a longer source, such as a journal article, and then exhibiting a writing burst in his draft.

In session 1, Baldur only opened the report template, quickly scrolled through it and shortly after ended the session, as evidenced by the absence of pauses, revisions and keystrokes (see Table 3 and 4). The use of the template, which included some demo text, explained the count of words in the initial text in this session (see Table 2). Baldur's writing process in session 2 was supported by three types of sources in the descending order of

time allocation: factual sources (mainly peer-reviewed articles) (appr. 8.5 minutes), course materials (mainly sample report and lecture materials on the structure of the report) (appr. 2.5 minutes) and language sources (appr. 1.5 minutes). Interestingly, most of the time spent on the sample report was just prior to formulation of the aim and research questions, which was then followed by writing meta comments about the wording in the research questions. In this session, the active process time amounted to 59% of the total session time, which indicates that 41% of the total time was available to high-level cognitive processes.<sup>1</sup> Session 2, which was devoted to producing text in the introduction chapter of the report, is characterised by the third highest increase in the text length (see Table 2) as well as the third highest number of revision events (see Table 3) and the third highest writing rate measured in keystrokes per minute (see Table 4). In session 3, Baldur chiefly focused on formulating and revising the heading and subheading of the report, during which the sample report was briefly consulted 7 times (in total 7 seconds). This short session is comprised of 68% of the active process time and 32% of the pause time and distinguishes itself by the second highest typing rate in keystrokes per minute and a relatively high number of revised characters.

The result unit (sessions 4–5) takes up more than half of the whole duration of the keylogging and amounts to about 7.5 hours of the total 11 hours (see Table 2). Consequently, sessions 4 and 5 are also the longest and the second longest respectively. The length of the session is also reflected in the increase in the text length, the number of pauses and revision events with session 4 topping the overall scores and session 5 following in the second place (see Tables 2, 3). In this unit, about 80% of the total number of words are produced, primarily in the result chapter but also in abstract, table of contents and discussion.

Session 4, being the longest session, indicated the increasing complexity of the process of academic writing. To prepare the result chapter, Baldur worked in total about 2 hours and 15 minutes with the Internet search to plan and generate ideas through accessing peer reviewed articles in such databases as PubMed, ScienceDirect and SpringerLink. As part of planning and evaluating his own writing, the student consulted the course materials, including the sample report and the lecture materials, which amounted to approximately 18 minutes. Baldur also spent approximately two and a half minutes on finding out standards of preparing an appendix as well as MS Word functions such as creating a table of contents. Further, the student used language sources to evaluate and revise the emerging text for approximately 5 minutes in total. These various types of sources were used continuously and multiple times in the writing process and hinted how the

---

<sup>1</sup> The percentages were calculated by the formula  $x/y*100$ , where x is active or pause time respectively (see Table 3) and y is time on task (see Table 2), and were rounded up or down to the nearest whole number.

three levels in the SVW-model interacted when the student appeared to be switching between planning, generating ideas, transcribing, evaluating and revising the text. As the result, this session showed the longest average pause length among all sessions (see Table 3) as well as typing and writing rate are among the lowest (see Table 4). It is striking though that Baldur spent only about 8 minutes on private and unrelated sources in this 306-minute-long session. The active process time in the session added up to 59% of the total time leaving 41% to the pauses. The complexity of the writing process was further increased due to the use of a separate Google document, in which Baldur stored and organised the references found on the Internet. To ease up the complexity, the student wrote several meta comments to monitor and guide his writing.

In session 5, Baldur exhibited a very different pattern of writing, which manifested itself in 80% of the total 136 minutes devoted to working in the report. Compared to the previous session, the student spent very little time on other sources while writing such as the course materials (appr. 5.5 minutes), factual sources (appr. 3 minutes) and language sources (appr. 1 minute). Thirteen minutes were reserved for private and unrelated activities. The sparse use of other sources contributed to the highest writing rate measured in keystrokes per minute in this session (see Table 4). At the end of the session, Baldur started working with the discussion chapter and demonstrated that he monitors and guides his writing by copying and pasting his research questions as well as the course instructions on discussion at the beginning of the chapter. To sum up, the active process time in session 5 amounted to 59% of the total session time, which indicates that 41% of the total time was available to high-level cognitive processes.

The discussion unit comprises sessions 6–8, which were chiefly used to write discussion and conclusion in the report and amounts to approximately 1 hour of the total logging time. In session 6, Baldur created one more version of the report in Word and mainly used it to save various references. The increased number of various versions of the report was a likely reason of an oversight on Baldur's part, when the student got confused and in session 8 opened the initial version of the report in session 7 instead of the final version (see Table 2). He effectively re-did revisions from session 7 in session 8.

In session 6, Baldur produced most of the text of the discussion and conclusion chapter, and his writing process was supported by three types of sources in the descending order of time allocation: factual sources (mainly peer-reviewed articles) (appr. 1.5 minutes), language sources (0.5 minutes) and course materials (appr. 0.5 minutes). The focus time on the factual sources was brief and indicated no close reading during the session. Language sources were primarily used to find language equivalents in the formal register, while course materials (mainly sample report and lecture materials on the structure of the report) were presumably used to assess

whether the necessary contents were covered in the text, which is indirectly supported by the added meta comment in the conclusion. In total, Baldur spent less than 4% of the total logging time on consulting the sources and made few revisions, chiefly at the end of each paragraph, which resulted in the third highest typing rate and the second highest writing rate in keystrokes per minute (see Tables 3, 4). While writing the discussion, he appeared to check his emerging text against the previously added meta comments, which he deleted as the text grew. Baldur also added meta comments about what to cover in the conclusion chapter for future use. All in all, session 6 was comprised of 68% of the active process time and 32% pause time of which 9% was reserved to informal chatting.

Sessions 7–8 were alike each other, during which the number of words in the report slightly decreased (see Table 2) as Baldur deleted the previously added meta comments in the conclusion chapter and added only one new sentence. The sources are scarcely used in these sessions apart from consulting the sample report for 7 seconds in session 7 and language sources to find synonyms for approximately 9 seconds in session 8. The only difference between the sessions was that in session 8 the student spent most of the time on moving and revising the added sentence in the conclusion. This increased the overall pause time from 36% in session 7 to 46% in session 8 and led to the third longest average pause length time in the latter (see Table 3). As already pointed out above, the data on revisions appear to be unreliable for these sessions.

The overall review unit includes sessions 9–12 and lasts over 40 minutes. In this unit Baldur focused on revisions through the whole report. The keylogging data reveal that at this stage the student had transferred the text to the LaTeX documents, a cloud version and a locally stored version. Some slight changes to the text had been made in LaTeX without logging it but Baldur, appearing to realise this already in session 10, started to log his writing in LaTeX and gradually transferred the text from the LaTeX documents to the Word document. In the last session, Baldur left the following comment: “Note for Inputlog: The previous copy and pasting, as well as in previous Inputlog recordings, was me editing things in LaTeX for formatting and copying them over. Hence the use of commands/brackets. I realise afterwards that I should’ve written in Word before (sic) LaTeX, and then copied, rather than the reverse and skewing the recordings of Inputlog”. Due to the text transfers, the writing rate in sessions 10–11 returned the two lowest values, while the typing rate did not drop equally dramatically as keyboard shortcuts were used frequently (see Table 4).

In session 9, which is the second shortest session and the last where the number of words in the report decreased, Baldur deleted parts of a sentence in the results chapter and opened LaTeX documents for the first time in the logging data. For no known reason, data on typing and writing rate are not available and the revision data appear to be unreliable (see Table 3 and 4).

In session 10, Baldur started to transfer text from the LaTeX document to various parts of the report in Word document. The logging data suggests that some copy-and-pasted text was partially different, and some was identical. The active process time in this session amounted to 76% leaving 24%, including 11% that were spent on informal chatting, to the pause time (see Table 3).

In session 11, Baldur switched between three versions of the report, one in Word and two in LaTeX. This session is the third longest session, in which 39% of the total session time was reserved to pauses (see Table 2, 3). The logging data suggest that most of the pause time was reserved to cognitive processes of re-reading the existing text, consulting language sources and revising the wordings. Switching between the documents and language revisions led to the third highest number of pauses and the second longest average pause length (see Table 3). The text was revised in all chapters of the report but introduction with most of the revisions being done in the discussion chapter. The new LaTeX document was chiefly used to store the final text that was then pasted into the Word document.

Session 12, which is the last logged session, was characterised by the lowest pause time of 19% that was available for high-level cognitive processes. In this session, Baldur copied the text from the LaTeX document to the Word document, although the text is primarily the same in both versions. This resulted in the highest number of revised characters and the highest typing rate in keystrokes per minute with a relatively low writing rate (see Tables 3, 4). The number of words in the report increased (see Table 2) as the reference posts were converted to the BibTeX system, after which Baldur deleted the meta comment about the converting. This suggested that the student finished the writing process and was not in any need to further monitor it.

## Discussion

To contribute to the understanding of academic writing processes, we used keystroke logging to explore an authentic academic writing process consisting of several, consecutive writing sessions focusing on the overall thematic structure of the writing process, on the writer's fluency, as measured by pausing and revision behaviours and production rate, as well as on the writer's digital source use.

As regards the first research question, the writing process lasted for 11 hours distributed over 12 sessions. The sessions were thematically divided into 4 units representing main topics in the process of writing the report: introductory, result, discussion and overall review. The units were largely well-defined with only minor overlap between units. The small incidence of overlap between the units can partially be attributed to the design of the writing task with intermediate submissions of drafts focusing on different parts of the report, such as the introduction with research questions, results

section and discussion. The design of the task, in turn, reflects the step-by-step organisation of the content in academic writing (Matsuhashi 1981). There is, however, some overlap between the units that may reflect the non-linear and recursive implementation of writing processes at the text production, transcription and monitoring level (Berninger & Amtmann 2003). Writers respond to aspects such as the text written so far, communicative goals, text sources, feedback from teachers and peers and so on, and engage in text production, transcription and monitoring non-linearly and recursively.

As regards the second and third research questions, our study indicated that the writer's pausing and revision behaviours, production rate and digital source use reflect the cognitive effort involved in academic writing. The writer needs to juggle cognitive processes to implement the specific language, genre, structural and formatting characteristics, to search for, understand and synthesise a large number of factual sources, and to critically evaluate different viewpoints, among other things (Matsuhashi 1981). This became particularly evident in the results unit (session 4 and 5), in which the bulk (approx. 80%) of the words remaining in the final text was produced. Session 4 and 5 showed different characteristics and foci in the writing process. Out of the 5 hours that comprised session 4, Baldur spent more than half of the time on consulting various sources, reflecting that academic writing is a reading-to-write task (e.g. Leijten et al. 2019; Plakans 2008). Two hours and 15 minutes were spent on idea generation by reading factual sources on the Internet, which suggests that Baldur created the content of the text *during* the writing session. Course materials, formatting sources and language sources were consulted primarily for planning, monitoring and revising purposes. The focus on planning at a macro-level by generating ideas for the topics to be covered in the results section was reflected in frequent pausing, the longest pauses among all sessions and a low typing and writing rate. This finding is in line with previous studies showing that macro-level planning of the conceptual and linguistic progression of a text affords longer pauses (e.g. Spelman Miller 2000, 2006). By contrast, in session 5, Baldur focused primarily on producing text. Out of 2 hours and 15 minutes, 80% of the time was spent on working with the report and only little time was spent on consulting different types of sources. The focus on producing text in this session was reflected in the highest writing rate among all sessions.

Considering the length of session 4 and 5 – 306 minutes and 136 minutes respectively – it is startling that there is very little downtime. Leijten et al. (2014) argued that downtime plays an important role for relieving fatigue and increasing concentration on the task at hand when solving longer and more complex writing tasks. Such downtime may reflect a writer's "meta-knowledge of their own motivational limits" (Leijten et al. 2014: 331). Baldur only spent 8 and 13 minutes on private and task-unrelated activities

during session 4 and 5 respectively. Also, there were no longer pauses indicating inactivity that could be interpreted as downtime, for example, going to the toilet or eating. A similar pattern was found in the other sessions. Rather, Baldur continually focused on solving the writing task and only marginally focused on other things while solving the task. We acknowledge the need for further research on downtime and other strategies for relieving fatigue and increasing concentration while solving longer and more complex writing tasks.

The current writing task involves writing for both informative and argumentative purposes (cf. Vandermeulen et al. 2020). As a literature review, the informative results section is based on reading, comparing and contrasting information from different sources in order to identify the overarching theme. The way in which Baldur uses sources when writing the results section is much in line with how Vandermeulen et al. (2020) describe an effective source use for informative writing. This is evidenced by frequent transitions between different sources in session 4 and few transitions between sources in session 5. Session 4 is devoted to reading, comparing and contrasting information in different sources to identify the topics and the overarching theme of the results section, while session 5 focuses more on integrating the information from the sources into the text.

Baldur also demonstrated an effective digital source use when writing for argumentative purposes in the introduction and discussion. An effective source use in argumentative writing is characterised by a considerable amount of time spent in sources and frequent transitions between sources and the text at the beginning of the writing process (Vandermeulen et al. 2020). Proficient writers initially focus on reading sources with the goal to support a position and on selecting information from sources and writing it down. When writing the introduction Baldur frequently switched between factual sources and the text. More specifically, he took a very quick look of several seconds at a longer source, such as a journal article, and then exhibited a writing burst in his draft.

Effective source use in argumentative writing is also characterised by little time spent in sources at the end of the writing process (Vandermeulen et al. 2020). Proficient writers focus on writing and revising the text rather than on selecting information from sources. In the discussion, Baldur followed this pattern. During session 6, in which most of the text for the discussion was produced, the focus time on the factual sources was brief and indicated no close reading of sources. Thus, Baldur exhibited several of the features that constitute an effective source use according to Vandermeulen et al. (2020). However, we recognise that an effective source use may take many forms and leave to future research to explore what may constitute effective source uses in different phases of academic writing.

Our study sheds light on the difficulties with assigning a writer a specific writing profile (cf. Galbraith 1999; Bowen & Van Waes 2020). The

combined pause and revision data indicated a flexible writing profile in Baldur depending on the task at hand and, arguably, on any approaching deadline for submission of a draft and receipt of feedback. We understand this flexibility as a strategy to deal with the specific complexities of writing an academic report (Matsuhashi 1981). For example, as the writing progressed step-by-step in three units focusing on the different parts of the report (introductory, result and discussion), there is no overall clear pause and revision pattern indicating a specific writing profile. In each unit Baldur first developed a draft of the introduction (session 2), result (session 4) and discussion (session 6) respectively. When developing these drafts Baldur made longer pauses than in the following sessions in each unit, which may indicate that Baldur is a high-self monitor who spends more time on initial planning and idea generation in each unit. On the other hand, Baldur also made more revisions in the first session than in the following sessions in each unit, which may indicate a low-self monitor who is using writing as a tool to sculpture the content of the text. The fourth unit (overall review) adds further complexity to the dynamics of Baldur's writing profile. We acknowledge the need for more research on the dynamics of writing profiles and more specifically on how writers solve different types of writing tasks by adjusting their writing profiles to the specific complexities of the writing task. Furthermore, as little is known about the benefits of being a high-self monitor or a low-self monitor (see Bowen & Van Waes 2020), future research also needs to explore how writing profile relates to text quality.

## Limitations

This case study based on rich keystroke logging data from several consecutive writing sessions has a few limitations. Given that the data were collected during an authentic writing task, Baldur had full control of the data collection and could choose when to start and stop logging and what to log. The procedure did not allow us to control for work in between the logged sessions, meaning that we do not fully know *if* Baldur worked with the text in the report between the logged sessions or if he read any (non)digital sources when he was not logging. Neither do we have exact knowledge of what information in the digital sources that caught his attention and how he integrated the selected information into the text he wrote. These seeming shortcomings could have been controlled for by asking Baldur to keep a diary on his entire work with the text, including his efforts outside Inputlog, and by using eye movement tracking to establish his focus while reading from the screen. However, these additional data collection methods would have undermined the integrity of the authentic writing process that we were aiming for, while there is some evidence that strengthen the option of excluding them. Firstly, the collected data shows that the number of words initially in each session coincides with the number of words at the end of the previous session, except in session 8, in which Baldur opened the initial

version from session 7 (see results section above). This indicates that Baldur did not actively work in the logged version of his report in between the sessions. Secondly, the references in the intermediate and final versions of Baldur's report do not contain printed books or other material than the digital sources that he consulted during the logged sessions. Finally, the comment that Baldur left in the last session about the text transfer as well as the keylogging data suggest that Inputlog remained relatively unnoticed until the last 40 minutes of the total 11 logged hours. This unobtrusive nature of Inputlog may justify its sole use in the current study, which strives for the authenticity of the writing process to be intact.

Despite the lack of other data collection methods, Baldur's reading can still be inferred from his observed behaviours during writing. In the case of digital source use, we found that Baldur frequently used various sources, especially in session 4 where he spent more than 2.5 hours consulting factual sources, course materials, language sources and formatting. This session was also characterised by a large number of pauses, the longest average pause length of all sessions and among the lowest writing and typing rate of all sessions. We interpret this as Baldur spending much time reading the sources that he is consulting to generate ideas and to define the task. Further, session 4 was also characterised by a large number of revisions. We take the frequent revisions to indicate that Baldur spent much time monitoring the writing process by reading to evaluate the text produced so far. Also, we observed that Baldur consulted course materials during writing to adjust the text to suit the specific requirements of the writing task. In this sense, Baldur engages in all three types of reading defined by Hayes (1996) during writing.

We suggest that future research combines keystroke logging with screen recordings, think-aloud protocols and/or interviews to explore in more detail qualitative (e.g. selection of information from sources, integration of information into the text) and quantitative (e.g. time spent consulting digital and non-digital sources at different stages of the writing process, transitions between different sources and between the text and different sources) aspects of source use during authentic academic writing.

Another limitation lies in some technological challenges of logging over an extended period of time. For some sessions, data were missing or proved to be unreliable for no obvious reasons. Writing and typing rate data were missing for session 9 (see Table 4), and the number of revision events as well as revised characters for sessions 7–9 proved to be unreliable after manual inspection of the log-files (see Table 3). This indicates limitations in the software and may potentially reduce the reliability of the other quantitative data reported in the paper. To increase the reliability, we have manually inspected the log-files to confirm the reliability of most of the data reported, which means the likelihood of reporting unreliable data in the

paper is small. We acknowledge the need for developing tools and manuals that can be reliably used for research-purposes.

## Conclusion

The current study showed that keystroke logging has the potential to extend existing research with new and valuable insights into authentic academic writing processes. Such insights may inform the teaching and learning of academic writing to effectively support university students' development of academic writing skills, thus underscoring the problem-oriented approach within educational linguistics. For example, knowledge about different steps in the writing process, the time allocation for solving a writing task of this magnitude as well as the time needed to complete different steps in the process may inform the design and progression of teaching. Baldur seems to have sufficient language and genre knowledge and does not need to spend a lot of time on the active writing of the report. Instead, he can focus on idea generation and spend a large proportion of the total process on consulting factual sources. The allocation of time for reading and understanding factual sources needs to be recognised in the teaching of academic writing skills. It is, however, important to consider that some students may not have sufficient language and genre knowledge and may need more support in how to plan and monitor their writing by consulting formatting sources, course materials and language sources. These students may also need additional time for planning and monitoring the writing. Also, the design of the writing task with intermediate submissions of drafts focusing on different parts of the report may support the writer in the writing process and facilitate the speed of text production. The findings indicated that Baldur wrote most efficiently and focused before deadlines for submitting drafts or after peer-review and feedback from the teacher. This highlights the importance of intermediate submissions for supporting students in the progression of the writing.

## References

- Alves, Rui A. & Limpo, Teresa (2015), "Progress in written language bursts, pauses, transcription, and written composition across schooling", *Scientific Studies of Reading* 19(5): 374–391.  
<https://doi.org/10.1080/10888438.2015.1059838>
- Berninger, W. Virginia & Amtmann, Dagmar (2003), "Preventing written expression disabilities through early and continuing assessment and intervention for handwriting and/or spelling problems: Research into practice", in R. Karen Harris and Steven Graham (Eds), *Handbook of learning disabilities*. New York: The Guilford Press, pp 345–363.
- Bowen, Neil & Van Waes, Luuk (2020), "Exploring revisions in academic text: closing the gap between process and product approaches in digital

- writing”, *Written Communication* 37(3): 322–364.  
<https://doi.org/10.1177/0741088320916508>
- Chenoweth, N. Ann & Hayes, John R. (2001), “Fluency in writing: Generating text in L1 and L2”, *Written Communication* 18(1): 80–98.  
<https://doi.org/10.1177/0741088301018001004>
- Chenu, Florence, Pellegrino, Francois, Jisa, Harriet & Fayol, Michel (2014), “Interword and intraword pause threshold in writing”, *Frontiers in Psychology* 5: 1–7. <https://doi.org/10.3389/fpsyg.2014.00182>
- Connelly, Vincent, Dockrell, Julie E., Walter, Kristy & Critten, Sarah (2012), “Predicting the quality of composition and written language bursts from oral language, spelling, and handwriting skills in children with and without specific language impairment”, *Written Communication* 29(3): 278–302. <https://doi.org/10.1177/0741088312451109>
- Cumming, Alistair, Kantor, Robert, Baba, Kyoko, Erdosy, Usman, Eouanqoui, Keanre & James, Mark (2005), “Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL”, *Assessing Writing* 10: 5–43. <https://doi.org/10.1016/j.asw.2005.02.001>
- Cumming, Alistair, Lai, Conttia & Cho, Hyeyoon (2016), “Students’ writing from sources for academic purposes: a synthesis of recent research”, *Journal of English for Academic Purposes* 23: 47–58.  
<https://doi.org/10.1016/j.jeap.2016.06.002>
- Davis, Mary (2013), “The development of source use by international postgraduate students”, *Journal of English for Academic Purposes* 12: 125–135. <https://doi.org/10.1016/j.jeap.2012.11.008>
- Escorcía, Dyanne, Passerault, Jean-Michel, Ros, Christine & Pylouster, Jean (2017), “Profiling writers: analysis of writing dynamics among college students”, *Metacognition Learning* 12: 233–273.  
<https://doi.org/10.1007/s11409-016-9166-6>
- Flower, Linda & Hayes, John R. (1981), “A cognitive process theory of writing”, *College Composition and Communication* 32(4): 365–387.  
<https://doi.org/10.2307/356600>
- Galbraith, David (1999), “Writing as a knowledge-constituting process”, in Mark Torrance and David Galbraith (Eds), *Knowing what to write: Conceptual processes in text production*. Amsterdam: Amsterdam University Press, pp. 137–157.
- Hayes, John R. (1996), “A new framework for understanding cognition and affect in writing”, in C. Michael Levy and Sarah Ransdell (Eds), *The science of writing: Theories, methods, individual differences, and applications*. Mahwah, New Jersey: Lawrence Erlbaum Associates, pp. 1–27.
- Hayes, John R. (2012), “Modeling and remodeling writing”, *Written Communication* 29(3): 369–388.  
<https://doi.org/10.1177/0741088312451260>
- Hyland, Theresa Ann (2009), “Drawing a line in the sand: identifying the borderzone between self and other in EL1 and EL2 citation practices”, *Assessing Writing* 14: 62–74. <https://doi.org/10.1016/j.asw.2009.01.001>

- Johansson, Roger, Wengelin, Åsa, Johansson, Victoria & Holmqvist, Kenneth (2010), "Looking at the keyboard or the monitor: relationship with text production processes", *Reading and Writing* 23: 835–851.  
<https://doi.org/10.1007/s11145-009-9189-3>
- Keck, Casey (2006), "The use of paraphrase in summary writing: a comparison of L1 and L2 writers", *Journal of Second Language Writing* 15: 261–278. <https://doi.org/10.1016/j.jslw.2006.09.006>
- Keck, Casey (2014), "Copying, paraphrasing, and academic writing development: a re-examination of L1 and L2 summarization practices", *Journal of Second Language Writing* 25: 4–22.  
<https://doi.org/10.1016/j.jslw.2014.05.005>
- Kellogg, Ronald T., Whiteford, Alison P., Turner, Casey E., Cahill, Michael & Mertens, Andrew (2013), "Working memory in written composition: an evaluation of the 1996 model", *Journal of Writing Research* 5(2): 159–190. <https://doi.org/10.17239/jowr-2013.05.02.1>
- Kobayashi, Hiroe & Rinnert, Carol (2013), "L1/L2/L3 writing development: Longitudinal case study of a Japanese multicompetent writer", *Journal of Second Language Writing* 22: 4–33.  
<https://doi.org/10.1016/j.jslw.2012.11.001>
- Leijten, Mariëlle & Van Waes, Luuk (2013), "Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes", *Written Communication* 30(3): 358–392.  
<https://doi.org/10.1177/0741088313491692>
- Leijten, Mariëlle, Van Waes, Luuk, Schriver, Karen & Hayes, John R. (2014), "Writing in the workplace: Construction documents using multiple digital sources", *Journal of Writing Research* 5(3): 285–337.  
<https://doi.org/10.17239/jowr-2014.05.03.3>
- Leijten, Mariëlle, Van Waes, Luuk, Schrijver, Iris, Bernolet, Sarah & Vangehuchten, Lieve (2019), "Mapping master's students' use of external sources in source-based writing in L1 and L2", *Studies in Second Language Acquisition* 41(3): 555–582.  
<https://doi.org/10.1017/S0272263119000251>
- Lindgren, Eva, Leijten, Mariëlle & Van Waes, Luuk (2011), "Adapting to the reader during writing", *Written Language & Literacy* 14(2): 188–223.  
<https://doi.org/10.1075/wll.14.2.02lin>
- Lindgren, Eva, Spelman Miller, Kristyan & Sullivan, Kirk P. H. (2008), "Development of fluency and revision in L1 and L2 writing in Swedish high school years eight and nine", *International Journal of Applied Linguistics* 156: 133–151.
- Lindgren, Eva & Sullivan, Kirk P. H. (2006), "Analysing online revision", in Kirk P. H. Sullivan and Eva Lindgren (Eds), *Computer Key-Stroke Logging and Writing: Methods and Applications*. Leiden: Brill, pp. 157–188.
- Lindgren, Eva & Sullivan, Kirk P. H. (Eds) (2019), *Observing writing: Insights from keystroke logging and handwriting*. Leiden: Brill.
- Liu, Gi-Zen, Lin, Vivien, Kou, Xiaojing & Wang, Hong-You (2016), "Best practices in L2 English source use pedagogy: a thematic review and

- synthesis of empirical studies”, *Educational Research Review* 19: 36–57.  
<https://doi.org/10.1016/j.edurev.2016.06.002>
- Mateos, Mar, Solé, Isabel, Martín, Elena, Cuevas, Isabel, Miras, Mariana & Castells, Nuria (2014), “Writing a synthesis from multiple sources as a learning activity”, in Perry Klein, Pietro Boscolo, Lori Kirkpatrick and Carmen Gelati (Eds), *Writing as a learning activity*. Leiden: Brill, pp. 169–190. [https://doi.org/10.1163/9789004265011\\_009](https://doi.org/10.1163/9789004265011_009)
- Matsuhashi, Ann (1981), “Pausing and planning: the tempo of written discourse production”, *Research in the Teaching of English* 15(2): 113–134.
- MacArthur, Charles A., Graham, Steve & Fitzgerald, John (Eds) (2008), *Handbook of writing research*. New York: The Guilford Press.
- McDonough, Kim, Crawford, William J. & De Vleeschauwer, Jindarat (2014), “Summary writing in a Thai EFL university context”, *Journal of Second Language Writing* 24: 20–32. <https://doi.org/10.1016/j.jslw.2014.03.001>
- Olive, Thierry & Kellogg, Ronald T. (2002), “Concurrent activation of high- and low-level production processes in written composition”, *Memory & Cognition* 30(4): 594–600.
- Olive, Thierry, Favart, Monik, Beauvais, Caroline & Beauvais, Lucie (2009), “Children’s cognitive effort and fluency in writing: effects of genre and of handwriting automatisation”, *Learning and Instruction* 19: 299–308.  
<https://doi.org/10.1016/j.learninstruc.2008.05.005>
- Plakans, Lia (2008), “Comparing composing processes in writing-only and reading-to-write test tasks”, *Assessing Writing* 13: 111–129.  
<https://doi.org/10.1016/j.asw.2008.07.001>
- Plakans, Lia (2009), “The role of reading strategies in integrated L2 writing tasks”, *Journal of English for Academic Purposes* 8: 252–266.  
<https://doi.org/10.1016/j.jeap.2009.05.001>
- Schilperoord, Joost (1996). *It’s about Time: Temporal Aspects of Cognitive Processes in Text Production*. Amsterdam: Rodopi.
- Schilperoord, Joost (2002), “On the cognitive status of pauses in discourse production”, in Thierry Olive and Michael C. Levy (Eds), *Contemporary Tools and Techniques for Studying Writing*. Dordrecht: Kluwer Academic Publishers, pp. 61–87. [https://doi.org/10.1007/978-94-010-0468-8\\_4](https://doi.org/10.1007/978-94-010-0468-8_4)
- Shi, Ling (2004), “Textual borrowing in second-language writing”, *Written Communication* 21(2): 171–200.  
<https://doi.org/10.1177/0741088303262846>
- Solé, Isabel, Miras, Mariana, Castells, Núria, Espino, Sandra & Minguela, Marta (2013), “Integrating information: an analysis of the processes involved and the products generated in a written synthesis task”, *Written Communication* 30(1): 63–90.  
<https://doi.org/10.1177/0741088312466532>
- Spelman Miller, Kristyan (2000), “Academic writers on-line: Investigating pausing in the production of text”, *Language Teaching Research* 4(2): 123–148. <https://doi.org/10.1177/13621688000400203>
- Spelman Miller, Kristyan (2006), “Pausing, productivity and the processing of topic in online writing”, in Kirk P. H. Sullivan and Eva Lindgren (Eds),

- Computer Key-Stroke Logging and Writing: Methods and Applications*. Leiden: Brill, pp. 131–155.
- Sumner, Emma, Connelly, Vincent & Barnett, Anna L. (2013), “Children with dyslexia are slow writers because they pause more often and not because they are slow at handwriting execution”, *Reading and Writing: An Interdisciplinary Journal* 26(6): 991–1008.  
<https://doi.org/10.1007/s11145-012-9403-6>
- Vandermeulen, Nina, van den Broek, Brenda, Van Steendam, Elke & Rijlaarsdam, Gert (2020), “In search of an effective source use pattern for writing argumentative and informative synthesis texts”, *Reading and Writing* 33: 239–266. <https://doi.org/10.1007/s11145-019-09958-3>
- Van Waes, Luuk & Leijten, Mariëlle (2015), “Fluency in writing: a multidimensional perspective on writing fluency applied to L1 and L2”, *Computers and Composition* 38: 79–95.  
<https://doi.org/10.1016/j.compcom.2015.09.012>
- Van Waes, Luuk & Schellens, Peter Jan (2003), “Writing profiles: the effect of the writing mode on pausing and revision patterns of experienced writers”, *Journal of Pragmatics* 35(6): 829–853.  
[https://doi.org/10.1016/S0378-2166\(02\)00121-2](https://doi.org/10.1016/S0378-2166(02)00121-2)
- Wengelin, Åsa (2006), “Examining pauses in writing: theory, methods and empirical data”, in Kirk P. H. Sullivan and Eva Lindgren (Eds), *Computer Key-Stroke Logging and Writing: Methods and Applications*. Leiden: Brill, pp. 107–130.
- Wengelin, Åsa, Johansson, Roger & Johansson, Victoria (2014), “Expressive writing in Swedish 15-year-olds with reading and writing difficulties”, in Barbara Arfé, Julie Dockrell and Virginia Berninger (Eds), *Writing Development in Children with Hearing Loss, Dyslexia or Oral Language Problems: Implications for Assessment and Instruction*. Oxford: Oxford University Press, pp. 244–256.