The effects of outlining on the writing process

A temporal analysis using keystroke logging

van Rijn, J.

Award date:
2021

Eindhoven, August 2021

The effects of outlining on the writing process

A temporal analysis using keystroke logging

Joris van Rijn
Identitiy number: 0950453

in partial fulfilment of the requirements for the degree of

Master of Science
in Human-Technology Interaction

Supervisors:
Dr. Ir. M.A. Conijn
Dr. U. Matzat

Abstract

Outlining is a popular writing strategy that helps a writer to plan the structure and content of a document prior to writing. While a large body of research has shown that outlining is related to the production of higher quality texts, much less is known about how this effect is achieved. The current study investigated how outlining affected the writing process, with the aim of getting a better understanding of how outlining works. The writing process was measured using keystroke logging, for which data was collected using a writing task in which 36 higher education students wrote an argumentative text with or without the instruction to create an outline beforehand. Multilevel regression analyses were performed to examine how the writing process differed between outlining conditions and how it varied over time. The results show that during the process of making an outline both writing fluency and revision behavior is reduced, whereas this is not the case during the process of writing the text. Moreover, outlining was also found to significantly affect pausing behaviors during writing. Overall, the study adds to earlier literature by showing that outlining could benefit writing by separating the cognitive processes involved in writing.

Contents

1 Introduction

These days, it is almost impossible to imagine a world without writing. Whether on paper in the form of a book, letter or newspaper or digitally using e-mails, blog posts and social media, it is safe to say that a substantial part of our communication relies on written words. Given that writing is such a valuable communication tool, it is not surprising that children are taught how to write from an early age, which is extended up to higher education. Having proper writing skills has been shown to be of major importance for achieving success both educationally and professionally (Kellogg & Whiteford, 2009). However, in the last decades, students' writing abilities have often been discussed as a matter of educational concern because many young people do not write well enough to meet the demands of school or the workplace (e.g. Graham & Perin, 2007; Lea & Street, 1998; Lillis & Turner, 2001).

With this problem in mind, researchers have tried to identify and analyze effective instructional practices to improve students' written communication skills. The results of these studies show that one of the best ways to enhance students' writing competencies, is to teach them writing strategies on how to plan, revise and edit their texts (Graham & Perin, 2007; Graham et al., 2012; Rogers & Graham, 2008). In particular outlining, a planning strategy that is used to organize one's ideas prior to writing, has consistently been shown to have beneficial effects on students' writing performance, leading to higher quality texts and a decrease in cognitive load (e.g. De Smet et al., 2014; Kellogg, 1988, 1990; Limpo & Alves, 2018; Isnard & Piolat, 1994). Most studies on outlining, however, have paid little attention to how outlining affects the writing process. Insights into how outlining affects the writing process, as opposed to the writing product, are important for two reasons. First, studies have shown that the writing process influences the quality of the written text (e.g. Baaijen & Galbraith, 2018; Sinharay et al., 2019). Since outlining is a writing strategy that is mainly targeted at the writing process (Kellogg, 1993), getting deeper insight into how it affects this process might reveal more about why outlining is such a beneficial strategy. Second, getting a better understanding of what influence outlining exerts on the writing process might provide us with a more detailed view of both the benefits and the limitations of this writing strategy. This knowledge can in turn be used to improve and develop writing tools and intervention methods to help students become better writers.

In writing research, common methods used to study the writing process are self-report measures (e.g. Limpo & Alves, 2018) and think-aloud protocols (e.g. Braaksma et al., 2004). However, as these are criticised for disrupting the writing process, key-stroke logging has been used more and more as an unobtrusive alternative (Leijten &

Van Waes, 2013). Keystroke logging comprises the recording of the keys struck on a keyboard, which results in a detailed overview of the real-time creation of a text. By analyzing these keystroke logs, insights can be derived about the underlying cognitive processes that are carried out during writing (Leijten & Van Waes, 2013; Strömqvist et al., 2006).

The current study aims at investigating the effects of outlining on the writing process by means of keystroke logging. Several studies have already used keystroke logging in the context of analyzing and evaluating a prewriting strategy such as outlining (e.g. Baaijen & Galbraith, 2018; Baaijen et al., 2014; De Smet et al., 2014). However, these studies mainly focused on global measures of the writing process and did not pay much attention to how the writing process was affected over time. Moreover, the studies that did try to take some temporal aspects into account, only did so in a minimal way, giving a rather narrow and incomplete view of the effects of outlining. Given this gap in the literature, the goal of the current study is to provide a broader view of the effects of outlining on the writing process.

2 Background literature

2.1 The writing process

Hayes & Flower's (1980) seminal model of writing provides the ground basis for the theoretical framework relevant to this study. In their model, they described the writing process in terms of three cognitive processes, namely, planning, translating, and reviewing. Here, planning involves the formulation of writing goals and the generation and organization of ideas, translating concerns the conversion of these ideas into written text, and reviewing describes the process of monitoring and editing the text that is written so far. An important aspect of the model by Hayes & Flower (1980) is that these three processes do not correspond to separate stages in the writing process but are applied in a recursive manner. Planning, translating, and reviewing processes constantly alternate and interact with each other and can be carried out at any moment during writing (Hayes & Flower, 1980).

Over the years, Hayes & Flower's (1980) model got adjusted a number of times (e.g. Hayes, 1996, 2012; Leijten et al., 2014). Most notable is Hayes' (2012) version of the model since it did not mention the cognitive processes of planning and revising in it anymore. These processes got removed as they were seen as specialized writing activities. Despite removing these processes Hayes (2012) recognized that this may have appeared counterintuitive as "we know that planning and revising happen" (p.375). The clear distinction Hayes & Flower's (1980) original model provides in describing the cognitive processes involved in writing (i.e. planning, translating and revising), in combination with the recognition that the terminology used in this model is still relevant, formed the most important reason to adopt this model to describe the theoretical base of the current study.

Studies have shown that coordinating the planning, translating and revising processes that take place during writing can be challenging, and may put great demands on the writers' working memory (Kellogg, 1996; Olive & Kellogg, 2002). Since working memory capacity is only limited and cognitive overload can have a negative impact on the writing performance, it is important for a writer to properly manage the cognitive load in order to create high quality texts (Beauvais et al., 2011; Olive, 2014). One way to do this is by reducing the overlap between the cognitive processes as this decreases the chances of them conflicting (Torrance & Galbraith, 2006). This can be done by means of writing strategies, such as outlining, that help the writer to divide a writing task into smaller subtasks (Torrance & Galbraith, 2006).

2.2 Writing strategies

Given the complexity of the writing process, splitting it up into smaller subtasks makes it a lot more manageable. These subtasks can then be performed in succession rather than concurrently. A writing strategy is the method in which a writer partitions and orders his or her writing process (Torrance et al., 1994). Writing strategies come in many shapes and sizes, but most of them can be categorized along two dimensions (Torrance et al., 2000). The first dimension describes the stage in which the content of the text is generated. Writers can either plan their content prior to writing (i.e. offline planning), or they can formulate their content during writing (i.e. online planning) (Galbraith, 1992). The second dimension describes the extent to which a writer reviews the text written so far. While some strategies involve extensive rereading and refining during writing, others aim at publishing a flawless text in one go (Torrance et al., 2000).

Writing strategies focused on planning prior to writing have been shown to work notably well for novice writers (Graham et al., 2012; Graham & Perin, 2007). This subclass of writing strategies is also referred to as prewriting strategies and they are specifically designed to support writers with generating or organizing ideas for their composition. Still, prewriting can take on many different forms, with some strategies focusing on specific types of writing, while others are more generic. Examples of prewriting strategies include listing, freewriting, outlining, clustering, and goal setting. While the effectiveness of each of these strategies is partly dependent on the writer and the writing task, outlining has been shown to improve writing performance among students at various ages and educational levels at a consistent basis (i.e. De La Paz & Graham, 2002; Ellis & Yuan, 2004; Galbraith et al., 2005; Galbraith & Torrance, 2004; Limpo & Alves, 2018; Page-Voth & Graham, 1999). What is less clear is why outlining generally tends to work so well. Therefore, the current study will take a closer look at outlining as a prewriting strategy.

2.3 Outlining

Outlining is a popular note-taking method that is used to plan the structure and contents of a document prior to writing (Hayes & Nash, 1996). An outline typically consists out of a hierarchically ordered list of topics and subtopics that is organized in the sequence which the writer intends to use for the final text. In general, an outline supports the writer with a space to easily consider ideas without having to write complete paragraphs or think about the coherence of the text. Hence, an outline is more than just a memory aid. The main benefit of outlining is that it helps writers to separate the planning and translation components of the writing process, which allows them to organize their ideas more effectively prior to writing, and in turn to focus their attention

more exclusively on translating ideas into words during the production of the text itself (Kellogg, 1988). This way, outlining could reduce the cognitive load experienced during writing which in turn can improve the quality of the text that is being written (De Smet et al., 2011, 2012; Galbraith & Rijlaarsdam, 1999; Kellogg, 1988, 1990).

Within an outline, a text can be planned out in different ways. Hayes & Nash (1996) created a framework that categorized the different types of planning described in the writing literature. This framework distinguished between two types of planning, 1) abstract planning, and 2) language planning. In abstract planning, one thinks of ideas without specifying the exact language that will be used whereas language planning already involves creating full sentences including correct grammar and syntax. Abstract planning can in turn be divided into two categories called content planning and non-content planning. Content planning is, like language planning, concerned with what the writer wants to say, but text is typically planned out in a simplified way for example by using keywords. Non-content planning on the other hand is concerned with aspects such as the structure, tone and style of the text. In an outline, non-content planning could include notes like "Explain the second argument" or "Make it sound professional". In general, outlines tend to include both notes related to content and notes related to non-content planning.

Studies have consistently shown that outline-based strategies have beneficial effects on students' writing performance (e.g. De La Paz & Graham, 2002; De Smet et al., 2011, 2014; Ellis & Yuan, 2004; Galbraith et al., 2005; Galbraith & Torrance, 2004; Kellogg, 1988, 1990; Limpo & Alves, 2018). Most of these studies focused on the effect of outlining on argumentative writing. For instance, Kellogg (1988; 1990) compared prewriting strategies in which students generated written outlines, mental outlines, cluster diagrams and rough full text drafts. His results showed that students who were asked to make an outline prior to writing, wrote higher quality essays than students in a control condition who were not instructed to make an outline. In particular, these students received higher ratings of idea development and were judged to write more effectively. Moreover, the results also showed that outlining was the superior prewriting strategy as clustering and drafting did not improve final product quality. The writing tasks in the studies by Kellogg (1988; 1990) were performed on paper. In a more recent set of studies, De Smet and colleagues (2011; 2014) conducted similar experiments but performed on the computer. They specifically investigated the effects of electronic outlining on writing, in which an outline had to be created using a digital tool. The studies found that making an outline resulted in higher product quality, including better text structure and more complex argumentation, compared to not making an outline. In addition, they found that the effects of outlining were stronger

with more practice, since students who outlined for a second time created better texts than students who outlined for the first time. Besides argumentative writing, Ellis & Yuan (2004) discovered that outlining also lead to improved text quality in narrative writing as it resulted in texts with higher grammatical complexity (measured by the range of different grammatical forms used). Overall, the findings in previous literature suggest a clear positive relationship between outlining and the quality of the writing product.

However, much less is known about what exactly causes these quality improvements as studies on outlining largely failed to pay detailed attention to the writing process. Moreover, the studies that did look at aspects of the writing process, still left certain aspects unexplored. For instance, several studies that examined the effect of outlining on writing quality, also investigated whether there was an effect on writing fluency (e.g. Kellogg, 1988, 1990; Ellis & Yuan, 2004). These studies mainly used global measures of writing fluency, such as the average number of words typed per minute (WPM). However, since writing is a dynamic process and fluency is likely to change during writing (Van Waes & Leijten, 2015), global measures could give a wrong impression of what actually happens during writing. Therefore, it is also important to consider fluency over time. Similarly, outlining could reduce revision behavior (Baaijen et al., 2014) which in the outlining literature is also only addressed with global measures. Another example is the study performed by Limpo & Alves (2018), who investigated the effects of outlining on writing dynamics. Writing dynamics is the term used to describe the recursiveness of the cognitive processes involved in writing (Olive et al., 2002). In their experiment, participants performed an argumentative writing task and had to repeatedly report their ongoing mental process (i.e. planning, translating, revising, and other) after hearing a beep sound. Afterwards, the number of occurrences for each process was counted and projected over time. The results showed some interesting patterns in terms of temporal organization. Students who made an outline namely engaged in more revision behavior towards the end of the writing process compared to students who were not asked to plan beforehand. Additionally, they found a clear decrease in planning behavior among the outliners towards the end of the writing task, whereas the non-outliners planned more continuously throughout the whole period of composing. Overall, this study nicely demonstrated how outlining can help to mentally separate the higher-level cognitive processes involved in writing. Nevertheless, there is still a gap between the occurrence of these processes and the lower-level behavioral actions (i.e. typing) performed that actually create the text. Understanding how outlining also affects lower-level text production processes is an important step in getting a deeper understanding of how and why outlining works. De Smet et al. (2014) touched on this

by analyzing outlining using keystroke logging, which is also the aim of the current study.

2.4 Analyzing the writing process: keystroke logging

Keystroke logging is a method that is used to collect keystroke presses during writing in real-time (Leijten & Van Waes, 2013). To do so, one must use a specialized keystroke logging software program. Every time a key is struck on the keyboard, the logging software records several properties including which key is pressed, the time the key is pressed, and the time the key is released. This result is a detailed dataset that depicts a writers' typing process (Strömqvist et al., 2006). The raw data cannot be immediately interpreted which is why during the analysis of these keystrokes, special features (i.e. variables) are created that better capture the information in the data (Leijten & Van Waes, 2013). Previous studies have created and used a wide variety of keystroke features to better understand the writing process. These keystroke features can roughly be categorized into four groups (Conijn, 2020). First, there are features related to pauses, such as the timings between words or sentences (e.g. Medimorec & Risko, 2017). Second are features relating to revision, such as the number of deletions (e.g. Conijn, Dux Speltz et al., 2020). Third are features related to fluency, for example the number of words typed per minute (e.g. Van Waes & Leijten, 2015). Lastly are features related to verbosity, such as the number of characters typed in a text (e.g. Allen et al., 2016).

Earlier studies have shown that keystroke features can provide information about the higher-level cognitive processes involved in writing. Features related to pauses, revisions and fluency are of main interest in this study as these are connected to outlining. Pauses between words and between sentences have been linked with planning processes (Baaijen et al., 2012; Medimorec & Risko, 2017). The duration of a pause is used as an indicator of mental effort, with longer pauses suggesting more effort (Van Waes et al., 2014). Furthermore, revision features have been associated with reviewing processes (Van Waes et al., 2014) and fluency features that describe a writer's production rate are linked to translating processes (Baaijen et al., 2012). As mentioned, a few studies have looked at keystroke features related to pauses, revisions and fluency in the context of outlining but mainly did so on a global level. Keystroke logging, however, also allows for analyzing such features from a temporal perspective.

De Smet et al. (2014) used keystroke logging to get temporal insights into how outlining affects pausing behaviors. In particular, they looked at the number of pauses and the median pause length. They split the writing process into five intervals of equal time length, and compared the pausing features between intervals and between outline

conditions (outlining vs. control group without outlining). In general, they found that pausing behaviors significantly varied over time, with more pauses in the beginning of writing and less pauses towards the end. However, no clear differences were found between outlining conditions regarding both the number of pauses and their duration. This could partly be explained by the fact that they defined a pause as a period of 2000 ms or more. This is a rather arbitrary threshold which does not take into account differences between individuals. Moreover, research has shown that important cognitive processes can also occur during pauses below 2000 ms (Van Waes & Leijten, 2015), and hence, these are not taken into account in this study. Another shortcoming of the study by De Smet et al. (2014) is that they did not consider revision behaviors and fluency from a temporal perspective.

Overall, our knowledge about how outlining affects the writing process is lacking depth as mostly global measures have been used to measure it. De Smet et al. (2014) made a first step in understanding the writing process on a more detailed level by looking at temporal patterns but did so only with respect to pausing behaviour. To get a broader view of how outlining affects the writing process, an analysis with a temporal approach is necessary which covers a wide range of writing features including measures of revisions, pauses and fluency. The current study will provide such analyses.

3 The current study

The literature shows that outlining as a prewriting strategy has beneficial effects on writing product quality. However, very little is currently known about how these effects are obtained. The current study therefore aims at investigating the effects of outlining on the writing process. Accordingly, this study aims to answer the main research question:

Research Question: *What is the effect of outlining on the writing process?*

Hereby, the writing process will be measured by means of keystroke logging, focusing on keystroke features that are commonly used in the writing analytics literature. More specifically, the current study will investigate three types of keystroke features, namely 1) features related revision behavior, 2) features related to pausing behavior, and 3) features related to writing fluency. Therefore, the main research question will be answered by means of the following subquestions.

Subquestion 1: *What is the effect of outlining on revision behavior during writing?*

Subquestion 2: *What is the effect of outlining on pausing behavior during writing?*

Subquestion 3: *What is the effect of outlining on writing fluency?*

Regarding revision behaviors, no clear hypothesis can be formed based on the current literature. On the one hand, Baaijen et al. (2014) have showed that outlining decreased overall revision behavior whereas on the other hand De Smet et al. (2014) did not find such a difference. Hence, the effects of outlining on revision behavior have to be explored. With respect to the second subquestion, based on the theory described, it is expected that pausing behavior will vary depending on the amount of (cognitive) planning that takes place. Hence, when one is making an outline it is expected that pauses will be longer as there is much planning involved. In turn, during the composition of the text, it is expected that outlining will lead to shorter pauses as the text is already largely planned out. While earlier work did not find an effect of outlining on pausing behavior (De Smet et al., 2014), the current study aims at using different, more detailed, measures to describe pausing behavior which yet might be able to show such patterns. A similar pattern is expected regarding the third subquestion. Outlining will cause an initial drop in fluency when one is planning the text, which subsequently induces an increase in fluency when one is composing it.

4 Method

4.1 Design

In the current study, the effect of outlining on the writing process was examined by means of an experiment in which participants had to write an argumentative essay. The experiment using a mixed experimental design with both a between-subjects factor and a within-subjects factor. The between-subjects factor was the writing instruction that the participants received prior to the writing task. Participants were randomly assigned to either the outline condition or the non-outline condition (i.e. the control condition). The within-subjects factor was time, which was manipulated by dividing the complete writing process into six five-minute intervals. Hence, outlining condition and time were the independent variables of this study. The dependent variable was the writing process, which was measured by means of keystroke logging. More specifically, the writing process was represented by means of nine keystroke features that were extracted from the keystroke data. These features covered aspects of writing fluency, verbosity, pausing behaviors and revision behaviors.

4.2 Participants

A random sample of 40 students was recruited for this study. The students were recruited using the local participant database of the Eindhoven University of Technology and participation was completely voluntary. The data of three participants was dropped from analysis because of technical deficiencies that occurred during the experiment. Additionally, the data of one participant was dropped because of not properly following the task instructions, leaving a final sample of 36 participants for analysis. From these participants, 13 subjects reported identifying as female and 23 as male. The average age was 23 (SD = 2.77). Participants were all higher education students, with 13 students doing a bachelor's degree, 22 studying for a master's degree and 1 studying for a PhD. Furthermore, all participants indicated to be fluent in English, with 35 students indicating English to be their second language and one indicating English to be the first language. Self-reported writing skills and writing experiences differed between participants, however these were evenly spread over the two conditions. Participation in this study was completely voluntary and participants received €7.50 as compensation.

Sample size justification

An a-priori power analysis was performed before the start of the experiment to ensure sufficient power (see Appendix A). This power analyses indicated the need for at least 68 participants. However, this desired amount of participants was not reached which is why the study design was changed into the form as presented in the current paper.

To still ensure this study had enough power, a post-hoc power analysis was performed which will be described in more detail here.

Relevant examples from the literature were considered to determine the expected effect sizes. Several studies had investigated the effects of different outlining strategies on writing, but these mainly focused on measures of fluency (Ellis & Yuan, 2004; Kellogg, 1990; Limpo & Alves, 2018). Ellis & Yuan (2004) were the only who mentioned effect sizes in their paper and reported a quite large effect size for fluency (d = 1.45). Although the studies by Kellogg (1990) and Limpo & Alves (2018) did not report effect sizes, they were similar in terms of design and sample size to the study by Ellis & Yuan (2004) and were also able to find significant differences in writing fluency between an outlining and a no-outlining condition. Based on these studies, the effect size of outlining on writing fluency was expected to be large. Not much research on the effects of outlining had focused on the other aspects of the writing process (i.e. pauses and revisions), and thus, the expected effect sizes here had to be estimated more conservatively than the one for fluency. Therefore, it was thought desirable to detect medium-to-large sizes ($f = 0.35$) for all features.

The post hoc power analysis (F-test - ANOVA repeated measures, within-between interaction) was performed using G*power (Faul et al., 2007) with a total sample size of 36 participants, two groups, six measurements and an alpha of 0.05. This analysis showed that the current study had a power of >0.99. Moreover, it showed that the current study was able to detect small effect sizes (minimum $f = 0.2$) with a power of 0.9, and thus, the power was considered to be adequate.

4.3 Writing task

A writing task was created for this experiment in which students were asked to write an argumentative essay (see Appendix B). An argumentative text was chosen as it is a relatively complex literary genre to write. In comparison to other types of text, such as for example a narrative, the translation process in argumentative writing cannot rely on simple structures such as causality, chronology, or spatial organization (Coirier et al., 2000). Therefore, outlining may be especially beneficial when writing an argumentative essay, as it could help the writer with translating their complicated mental structures of arguments into a linear text.

The students were asked to write an argumentative essay about a current, much discussed topic, namely the use of vaccine passports during the COVID-19 pandemic. This topic was chosen for two reasons. Firstly, students had to be somewhat familiar with the topic as outlining is not beneficial when an author struggles with generating ideas (Kellogg, 1988). On the other hand, it was undesirable to provide too much

context information to the students as this could stimulate idea generation which in turn could also decrease the effect of outlining. Since everybody was familiar with the COVID-19 pandemic, and the use of vaccine passports was widely discussed in news channels and talk shows, the given topic fulfilled this requirement. Secondly, the topic had to be interesting for the students. Interest in the topic could increase the motivation for students to write which made it more likely they would take the task seriously and do their best. The chosen topic was thought to be fitting in this case, as social events are highly valued by students.

4.4 Procedure

The experiment was carried out on a laptop and took place in a controlled lab environment at the campus of the TU/e. Participants were first asked to complete an informed consent form, after which they could start with the experiment. Participants were randomly assigned to either the outline or the non-outline condition. Corresponding to the condition, participants first received a general instruction with the procedure of the study. These instructions provided information about the kind of task that had to be performed and its duration. Since participation was voluntary and the essays were not part of a study program, the general instructions also emphasized that it was important to take the task seriously and that it was not allowed to use any help sources such as a mobile phone or the internet. Thereafter, participants were given the writing task.

The students received 30 minutes for the writing task for which they had to write an essay of at least 200 words long. The essay had to be written in English. The amount of 200 words was chosen as this was considered to be the minimum number of words necessary to conduct an insightful keystroke analysis. For the participants in the outline condition, the writing task was split up into a 10-minute period to create the outline and a 20-minute period to write the essay. This division of outlining and writing time was also used in earlier work on the effects of planning which reported significant results (Ellis & Yuan, 2004; Limpo & Alves, 2018). The outline participants received specific instructions on how the outline had to be created (see appendix C) and were asked to work on the outline for the full 10 minutes that were given. When the 10 minutes were over, they were asked to stop outlining and instructed to start writing their essay for which they had 20 minutes left. The participants who were not required to create an outline started immediately with writing their essay and could work continuously for 30 minutes. Participants were allowed to stop earlier when they completed the task before the 30 minutes had expired. After finishing the writing task, the participants were asked to complete a post-task survey. When this survey was completed, participants were debriefed, compensated, and thanked for their

participation. Overall, the experiment had an approximate duration of 45 minutes.

The writing task, the post-task survey and all instructions were collectively implemented into one large survey using the online survey tool LimeSurvey. During the writing task, this survey displayed timers at the bottom of the screen such that participants could see how much time they had left. These timers also assured that participants would not take more time than was allowed, as they automatically directed participants to the next step in the experiment when time had run out. Moreover, in an opposite fashion, the timers assured that participants in the outline condition took the full 10 minutes to create their outline. For the duration of the whole experiment, the survey was shown on the right side of the laptop screen such that instructions and timers were always visible. On the left side of the screen was a Word document in which the participants could type their text. The keystroke data was collected from this Word document using the keystroke logging software Inputlog (Leijten & Van Waes, 2013).

4.5 Measurements

As mentioned, participants completed a post-task survey that focused on measuring several personal characteristics. Firstly, participants were asked about their preferred writing style. This was measured by means of a questionnaire developed by Kieft et al. (2008) that consists out of two independent scales that respectively describe the extent to which one usually tends to engage in planning behaviors (11 items) and revising behaviors (15 items) (see Appendix D for the exact items). These scales were assessed using a five-point Likert scale ranging from strongly disagree to strongly agree. A final measure was created by averaging the participants' answers for each scale. For the planning scale, the original set of 11 items showed low reliability ($\alpha = 0.667$). In order to improve the scale, the items were evaluated to check if they were all relevant. Three items were found to be formulated rather ambiguously or were not really typical for planning behavior. To elaborate on one example: *"When I reread my texts, sometimes they are very chaotic"*. This item was not considered to describe planning behavior accurately, as one would expect that a planned out text is less chaotic. Moreover, the term 'sometimes' made the statement vague, as certainly anybody could 'sometimes' perceive their text to be chaotic. Therefore, these three items were dropped from the scale which made it more reliable ($\alpha = 0.741$) and suitable for analysis. Regarding the revising scale, the measure was very unreliable ($\alpha = 0.487$). As this measure could not be improved, and the focus of this research was more specifically aimed at investigating planning behaviors, it was decided to not use this measure for further analyses.

Secondly, a number of self-report items were included to measure the participants' writing skills and experience (see Appendix E). These asked about how good

they considered their academic writing skills to be, how often they write per month, and whether they had significant writing experiences such as for example from side jobs, or a writing course.

Lastly, a set of demographic questions was included that elicited information on the participants' age, gender, educational level, native language, and non-native fluent languages.

4.6 Pre-processing and feature extraction

Before feature extraction, it was examined whether participants did not use more time than the 30 minutes that were assigned for the writing task. This was done by comparing the timestamps made by the survey at the start and end of the timers, with the raw keystroke data. Several participants appeared to have worked on their writings for 1-2 minutes longer than allowed. Replays of the writing process revealed that these participants often made small notes before starting with writing, and took some extra time finishing their sentence after the timer ended. To keep the time on task constant, these extra keystrokes at the beginning and end of the process were trimmed off the data and left out of the analysis. Thereafter, the keystroke features were extracted.

The keystroke features were extracted from the trimmed data using the software package R (2021). Methods and R code for the keystroke feature extraction were largely adopted from Conijn, Cook et al. (2020). In total, nine features were extracted from the keystroke log. The keystrokes can be categorized in features related to revision behavior, features related to pausing behavior, features related to writing fluency and features related to verbosity.

Features related to revisions. The goal here was to get a broader view of the effect of outlining on revision behavior. Hence, not only the amount of revisions was of interest, but also the properties of these revisions. Revisions can be categorized according to many properties (Conijn, Dux Speltz et al., 2020) but for the current study it was decided to focus on two in particular, their linguistic domain (i.e. their size) and their spatial location (i.e. where the revision occured in the text). The following three keystroke features related to revision behavior were extracted:

- *Number of revisions.* Number of insertions away from the point of inscription plus the number of sequences of backspaces and delete keystrokes (Barkaoui, 2016).
- *Ratio intext revisions.* Number of revisions away from the leading edge, divided by the total number of revisions. A higher ratio indicated that more revisions take place within the text that was already written, than at the point of inscription.
- *Number of backspaces per revision.* The number of backspaces divided by the number of revisions. A higher number of backspaces per revision indicate larger

revisions.

Features related to pauses. Pauses have been related to higher-level cognitive processes (Medimorec & Risko, 2017; Baaijen et al., 2012). Moreover, the location and duration of pauses could reveal something about the nature of these underlying cognitive processes (Medimorec & Risko, 2017). Therefore, three features were included to capture pausing behaviors during the writing process. All three of these features were based on the transition times between subsequent keystrokes (i.e. the interkeystroke intervals (IKI)). The distributions of these IKI's were often skewed to the right. Therefore, in line with the approach used by (Conijn, Cook et al., 2020), these features were log-transformed and all values above the 95th percentile were removed.

- *Mean time between words.* Time from the key press of the last letter of a word until the key press of the first letter of the next word (measured in milliseconds) (Sinharay et al., 2019; Chukharev-Hudilainen et al., 2019).
- *Mean time between sentences.* Time from the key press of the end of a sentence marker until the key press of the first letter of the next sentence (measured in milliseconds) (Baaijen & Galbraith, 2018).
- *Ratio long pauses between words.* Number of long pauses between words, divided by the total number of pauses between words (Baaijen & Galbraith, 2018). In the current study, a long pause is defined as a pause longer than two SD from the mean IKI.

Features related to writing fluency. Two keystroke features were included to measure writing fluency. These features were adopted from Van Waes & Leijten (2015) who, by means of principal component analyses, showed that these features describe the fluency of text production.

- *Number of characters per minute.* The number of characters typed during the process, divided by the time in minutes.
- *Ratio characters typed in P-bursts.* The number of characters typed in production bursts (P-bursts), divided by the total number of characters. A P-burst is defined as the set of actions between two long pauses.

Features related to verbosity Lastly, a general feature related to verbosity was included. While this feature is of less interest from a theoretical point of view, it still provides useful information about the amount of activity that is happening on the keyboard.

- *Number of keystroke events.* The number of keyboard events during the writing process. This includes character keys, navigation keys, as well as non-character keys such as Backspace, Delete and Enter.

As this study focused on the temporal organization of writing activities, the writing process was divided into 5-minute time intervals. Hence, every keystroke feature was calculated for each time interval resulting in a maximum of six datapoints per participant. The choice for using 5-minute intervals was based on the 10-minute period participants in the outline condition had to create their outline. This way, an easy distinction could be made between text written in the outline-phase and the essay-phase. Using 10-minute intervals was also an option, but as this results in fewer data points, this would lead to less detailed temporal insights and lower power for the analyses, which is why this was not preferred. Since participants were allowed to stop earlier when they were finished, not all 36 participants had six datapoints. In the final dataset, the first three time intervals (0-5, 5-10 and 10-15 minutes) had 36 datapoints (i.e. these were complete), the fourth time interval (15-20 minutes) had 34 datapoints, the fifth interval (20-25 minutes) had 31 datapoints, and the last time interval (25-30 minutes) had 28 datapoints. Hence, the final analysis was performed on a data set of 201 observations.

4.7 Statistical analysis

Data analysis was performed using the statistical software package Stata. To examine the effect of outlining on the writing process, a series of multilevel regression analyses were performed. A separate random intercept model was created to predict each of the nine keystroke features. For these models, the data was clustered by participant to take individual differences into account. The need for clustering on individual level was determined based on 'null' models without predictor variables that were created prior to creating the full models. In the full models, condition, time interval and the interaction effect between condition and time interval were included as the main predictor variables. Here, condition was represented with a binary variable in which a 1 meant that one was in the outline condition. Time intervals were coded with the numbers 1-6 for respectively the first to sixth time interval. Additionally, the measure for planning writing strategy was added to the models as control variable. Since, planning strategy did not significantly explain variance for any of the features, it was eventually excluded in all analyses.

For all multilevel models, assumptions were checked, and violations were found for either normality of residuals or homoscedasticity in all cases. Besides the time-based variables that were already log-transformed, further transformations of any feature could not improve the models. As this was expected due to the inherent noisiness of keystroke data, robust regressions were performed for all features.

Additional to the regression analysis, graphs were created to visualize the

temporal patterns of each feature. This way, deeper insights could be gained about how each feature changes with respect to time.

5 Results

This section is divided into four parts. In the first part I will present several descriptive statistics. In the three parts that follow, I will present the results of the multilevel regression analyses.

5.1 Descriptive statistics

From the total of 36 participants, 17 were assigned to the outline-condition and 19 to the control condition. Participants spent on average 26.7 minutes (SD=5.1) on the writing task. Participants in the control condition worked on average 24.6 minutes on the task (SD=5.8), while participants in the outline condition spent a little longer with an average of 29.3 minutes (SD=2.7). Moreover, the average essay length was 456 words long (SD=164.4), with outliners (M=468.7 words, SD=146.2) having slightly longer essays than the non-outliners (M=444.3 words, SD=149.6). Descriptive statistics of the keystroke features were calculated over the complete writing process and are displayed in Table 1.

	Control condition		**Outline condition**		**Total**	
Keystroke feature	Mean	SD	Mean	SD	Mean	SD
Number of revisions	94.7	45.5	91.9	49.8	93.4	46.7
Percentage of intext revisions	0.38	0.29	0.64	0.35	0.51	0.34
Number of backspaces per revision	7.9	8.5	8.4	5.8	8.2	7.3
Mean time between words (ms)	592.5	201.4	534.7	165.4	563.3	185.9
Mean time between sentences (ms)	2132.2	2222.6	1521.8	1103.3	1838.4	1795.3
Ratio long pauses between word	0.073	0.035	0.067	0.031	0.070	0.033
Number of characters per minute	164.5	38.2	162.1	34.8	163.4	36.1
Percentage of characters in P-bursts	0.23	0.29	0.13	0.23	0.18	0.27
Number of keystroke events	4013.8	1361.4	4697.9	939.0	4336.9	1214.98

Table 1: Descriptive statistics of each keystroke feature calculated per condition.

5.2 Features related to revision

To assess revision behavior, three features were analyzed: (1) the number of revisions, (2) the ratio in-text revisions, and (3) the number of backspaces per revision. In figure 1A-C, the revision features are visualized over time per condition. Here, it can clearly be seen that for the number of revisions and the ratio in-text revisions the differences between conditions are largest in the first 3 time intervals. The graph displaying the number of backspaces per revision indicates that this feature is more similar over the conditions but still shows quite some variations over time.

The results of the multilevel regression analyses of the revision features are presented in table 2. The models of all three features show no significant main effect

of outlining. This means that outlining does not have an effect on these features when looking over the complete writing process (i.e. averaged over time intervals). Nevertheless, all models do show a significant main effect of time. Moreover, significant interactions between outlining and time interval were found for both the number of revisions and the ratio in-text revisions. This means that the variations of these two features over time are influenced by whether one makes an outline or not. The interaction effects were found for the first three measurement periods which are also the stages where the outlining took place. Hence, in line with the graphs, these results indicate that outlining mainly affected these features during outlining. No significant interaction effects between outlining and time interval were found for the ratio backspaces per revision. Thus, while the ratio backspaces per revision did vary over time, this variation was not significantly influenced by the condition.

The intraclass-correlation coefficient (ICC) was calculated for each model to investigate the amount of variance explained by differences between individuals. Among the revision features, the largest ICC value was found for the ratio of in-text revisions as 43.2% of the variance in the ratio in-text revisions could be explained by individual differences.

	Nr. of revisions	**Ratio intext revisions**	**Number of backspaces per revision**
Outline	4.11 (4.28)	-0.04 (0.13)	2.31 (5.53)
Time interval	-2.77 (0.92) **	0.10 (0.03)***	0.84 (0.33)*
Interactions outline X time interval			
Outline X interval 1	-13.56 (6.13)*	0.24 (0.18)	-6.05 (5.29)
Outline X interval 2	-11.08 (5.52)*	0.43 (0.15)**	-3.03 (4.87)
Outline X interval 3	-6.19 (5.06)	0.28 (0.12)*	0.46 (6.35)
Outline X interval 4	-2.48 (4.51)	0.19 (0.10)	1.82 (3.98)
Outline X interval 5	- 0.35 (5.39)	0.09 (0.09)	-1.53 (3.89)
Outline X interval 6	*Reference*	*Reference*	*Reference*
Constants	27.10 (3.67)***	0.24 (0.11)*	7.11 (2.18)**
ICC	0.246	0.432	0.379

Table 2: Results of the multilevel regression analyses of the features related to revision. In each cell, the first number represents the regression coefficient β. The second number (in parentheses) represents the robust standard error. All interaction effects with time interval are relative to the sixth time interval. *p<0.05, **p<0.01, ***p<0.001

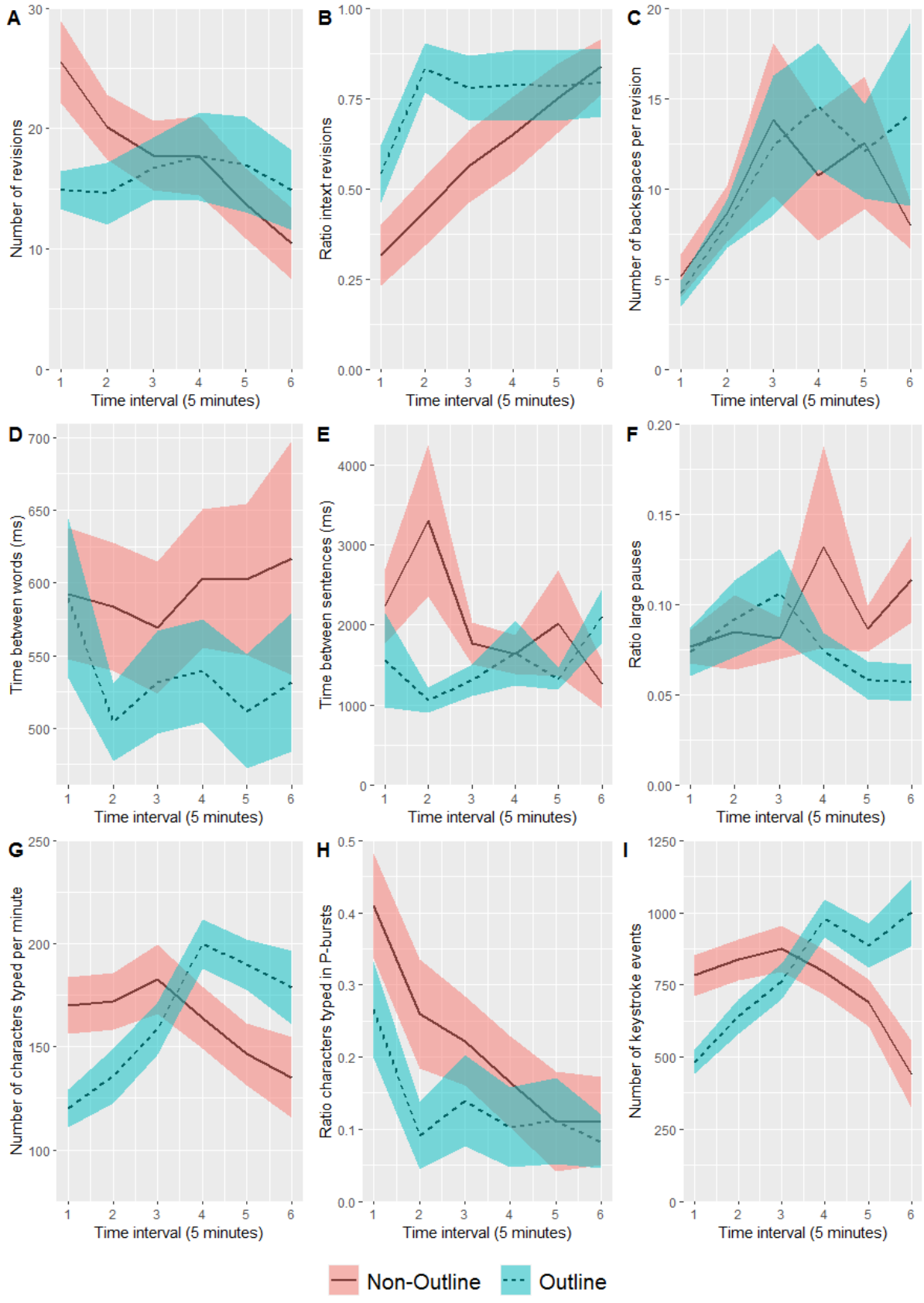

Figure 1: Visualization of the temporal variations of the keystroke features. Graphs A-C (top row) represent the features related to revision, graphs D-F represent the features related to pauses (middle row), and graphs G-I represent the features related to fluency and verbosity (bottom row). The colored ribbons (in red and blue) indicate the standard errors per condition.

5.3 Features related to pauses

In the same manner as with the features related to revision, multilevel regression analyses were performed to investigate the effect of outlining on the features related to pauses. The results of these analyses are presented in table 3. Figure 1D-F displays these features for both conditions plotted per time interval. The model for the mean time between words did not show any significant main effects or interaction effects of condition and time interval. This indicates that neither of these two variables had an effect on the amount of pausing between words. Both the model and the graph (figure 1D) do suggest that the time between words is generally shorter for participants in the outline condition, but these results evidently failed to reach statistical significance. Nevertheless, significant main and interaction effects were found for the other two features in this category.

Regarding the mean time between sentences, the results indicated a significant main effect of time, whereas the effect of outlining turned out to be not significant. Additionally, significant interaction effects were found for this feature between outlining and all five time intervals. Together, these results demonstrate that the mean time between sentences is, especially in the early stages of the writing process, significantly shorter when one is asked to create an outline prior to writing. This can also be seen in figure 1E.

For the ratio long pauses between words, the results indicated a significant main effect of outlining but not for time. Furthermore, the results showed significant interaction effects between outlining and the first four time periods. Interestingly, as can be seen in figure 1F, the ratio long pauses among outliners is higher in the first three time intervals, but lower in the fourth time interval. This indicates that devoting more attention to planning at the start of writing could in turn make writing less effortful towards the end of the writing process.

The ICC values point out that the proportion of variance explained by individual differences vary a lot between the pause features. What stands out is the high ICC of the mean time between words, signifying that according to this model, 85.1% of the variance in this feature can be explained by differences between participants. This means that most of the variations observed in this feature can be allocated to personal characteristics.

	Mean time between words	**Mean time between sentences**	**Ratio long pauses between words**
Outline	-0.038 (0.099)	0.45 (0.24)	-0.06 (0.03)*
Time interval	-0.003 (0.007)	-0.06 (0.03)*	0.01 (0.01)
Interactions outline X time interval			
Outline X interval 1	0.073 (0.061)	-0.78 (0.30)**	0.06 (0.03)*
Outline X interval 2	-0.021 (0.051)	-0.72 (0.25)**	0.07 (0.03)*
Outline X interval 3	0.009 (0.049)	-0.58 (0.20)**	0.07 (0.03)*
Outline X interval 4	0.012 (0.034)	-0.51 (0.24)*	0.03 (0.02)*
Outline X interval 5	-0.011 (0.026)	-0.36 (0.16)*	0.01 (0.01)
Outline X interval 6	*Reference*	*Reference*	*Reference*
Constants	6.054 (0.060)***	7.01 (0.16)***	0.07 (0.01)*
ICC	0.856	0.376	0.234

Table 3: Results of the multilevel regression analyses of the features related to pauses. In each cell, the first number represents the regression coefficient β. The second number (in parentheses) represents the robust standard error. All interaction effects with time interval are relative to the sixth time interval. *p<0.05, **p<0.01, ***p<0.001

5.4 Features related to fluency and verbosity

Lastly, the results of the multilevel analysis for the features related to fluency and verbosity are displayed in table 4. Visualisations of these features over time are presented in figure 1G-I. Starting with the two fluency features, the number of characters typed per minute and the ratio characters typed in P-bursts, the results indicate no main effect of outlining, but do indicate a main effect of time. Hence, for both fluency features there were significant differences between the repeated measures but not between conditions.

In addition, when looking at the number of characters per minute, significant negative interactions were found for the first two time intervals. As can be seen in figure 1G, this indicates that fluency is decreased when one is creating an outlining compared to when one starts writing directly. However, even though not significant, the results also indicate that outliners produce a higher number of characters per minute than the non-outliners during the last three time intervals. Overall, these results suggest that outlining itself decreases fluency, but in turn, induces an increase in fluency for the writing of the text that follows. Regarding the ratio characters in P-bursts, a significant negative interaction effect was only found for the second time interval. Thus, similar to the other fluency feature, this means that fluency is decreased during the second half of the outlining phase. Remarkably, figure 1H shows that over the duration of the complete writing process the ratio characters produced in P-bursts is lower for the participants in the outline condition. Thus, a comparison of the results of both fluency features suggest that increase an in one, does not necessarily correspond to an increase in the other.

Turning to the verbosity feature, the number of keystroke events, the results revealed significant main effects of both outlining and time. In addition, significant interaction effects were found between outlining and the first three time intervals. These results are very resembling to those of the number of characters typed per minute feature. In general, these outcomes demonstrate that the number of keystrokes were significantly lower for outline-subjects in the first half of the writing task (see figure 1I).

	Nr. of characters per minute	**Ratio characters in P-bursts**	**Nr. of keystroke events**
Outline	34.76 (22.03)	0.01 (0.07)	407.99 (138.38)**
Time interval	-6.88 (2.91)*	-0.06 (0.02)**	-57.45 (18.82)**
Interactions outline X time interval			
Outline X interval 1	-94.93 (23.80)***	-0.10 (0.12)	-811.76 (152.00)***
Outline X interval 2	-72.39 (28.25)*	-0.22 (0.09)*	-596.37 (161.27)***
Outline X interval 3	-42.36 (22.30)	-0.11 (0.09)	-416.62 (134.53)**
Outline X interval 4	5.57 (20.15)	-0.09 (0.07)	-142.29 (126.37)
Outline X interval 5	2.34 (19.05)	-0.03 (0.07)	-179.25 (132.55)
Outline X interval 6	*Reference*	*Reference*	*Reference*
Constants	187.16 (11.87)***	0.41 (0.08)***	943.93 (67.56)***
ICC	0.246	0.276	0.192

Table 4: Results of the multilevel regression analyses of the features related to fluency and verbosity. In each cell, the first number represents the regression coefficient β. The second number (in parentheses) represents the robust standard error. All interaction effects with time interval are relative to the sixth time interval. *p<0.05, **p<0.01, ***p<0.001

6 Discussion

The aim of the current study was to investigate how outlining affects the writing process. To do this, process measures related to revisions, pauses and writing fluency were derived from the writing process using keystroke logging and compared between two outline conditions.

6.1 The effect of outlining on revision behavior

The first sub question of this study sought to determine the effect of outlining on revision behavior. The results of the multilevel regression analyses revealed that outlining did not directly affect any of the features related to revision behavior. That is, no main effects of outlining were found on the number of revisions, the ratio intext revisions, and the ratio backspaces per revision. On the contrary, the results indicated significant variations over time for all three features .

Regarding the number of revisions, the results also showed that outlining lead to fewer revisions during the first two time intervals (i.e. during the outlining period), but that this was not the case during the rest of the writing process. Earlier work had suggested that outlining reduces the amount of revising during writing (Baaijen et al., 2014) but this hypothesis was not replicated (De Smet et al., 2014). The results of the current study also suggest that outlining does not affect the overall amount of revising but shows that there are differences between the outlining phase and the writing phase. The findings could be explained by the nature of the manipulation. Since generating the content of the work is an important part of creating the outline, the writers' main concern during the outlining phase is getting ideas on paper. As long as these ideas are not fully translated into a text, there is also less need for revising.

Looking at the ratio of intext revisions, the results again showed that outlining only had an effect during the first half of the writing process since participants in the outline conditions made more intext revisions than participants in the control condition during these time intervals. This observed increase in intext revisions could be attributed to an inherently higher focus on the structure and organization of the text. That is, when one is creating an outline, one is likely to go back in the outline to for example add new ideas or change the order of elements. In addition, the visualization of this feature showed that participants in the outline condition had a relatively constant rate of intext revisions when writing the essay, whereas the control group had more of a linear increase. As outlines could be used to impose order and hierarchy on the writers' ideas (Walvoord et al., 1995), these findings may partly be explained by a continuous reflection process in which outliners regularly compare their written text with their outline on whether they correspond. Non-outliners don't have a a predeter-

mined structure to hold on to which is why intext revisions might be more postponed towards the end, when they have a clearer view of what their text is going to be like. Nevertheless, it must be noted that these findings may be somewhat limited by the method this feature is extracted from the data. An intext revision is namely defined as a revision that did not take place at the point of inscription. The disadvantage of this definition is that when a student makes a small note on the bottom of the page and writes the entire essay before this note, all revisions are defined as intext revisions. This could in turn have a large effect on the results as no distinction can be made between actual intext revisions and extensions of the text that take place at the point of inscription (Lindgren et al., 2019).

In the same vein as the number of revisions and ratio intext revisions, the ratio backspaces per revisions was not affected by the manipulation. Significant differences were found over time, but in contrast to the other revision features, there were no significant interaction effects. This feature was included to investigate whether outlining influenced the relative size of revisions. In general, the results suggest that participants made more revisions of smaller pieces of text, indicating that there was a higher focus on low-level revisions (Barkaoui, 2016). This result may be explained by the fact that almost all participants wrote in their second language (L2). As L2 writers are more likely to make linguistic mistakes, these mistakes could tempt the writer to focus more on linguistic revisions (Broekkamp & Van Den Bergh, 1996).

6.2 The effect of outlining on pausing behavior

The objective of the second sub question was to identify how outlining affected features related to pausing behavior. No effects of outlining were found for the mean time between words, making it the only feature for which not a single effect was found. This was in line with earlier work which also failed to find an affect of outlining on the duration of pauses (De Smet et al., 2014). This finding may be explained by the fact that individual differences play a large role here, as according to the multilevel regression model, 85.1% of the differences in the mean time between words could be explained by differences between individuals. In general, the time between words has been used a lot in the writing analytics literature. For instance, studies have shown that pause timings between words could be used as a predictor for writing quality (Conijn, Cook et al., 2020), and that they could be affected by differences in task complexity, cognitive load (Conijn et al., 2019) or whether one types in their first or second language (Chukharev-Hudilainen et al., 2019). Hence, a difference in the mean time between words could mean multiple things suggesting that this feature does not only reflect planning processes, and thus, this feature should be interpreted with a bit

more caution.

Turning to the next feature, the results showed that mean time between sentences was longer for the subjects in the non-outline group during all time intervals. This was in line with the expectations which stated that outlining would lead to shorter pauses during writing. What is interesting however, is that outliners on average also had a shorter mean time between sentences during the first two time intervals were the outlining took place. As most of the planning was expected to take place during the ten minutes of outlining, it was expected that this would lead to longer pausing times. These results may be explained by how an outline is created. To explain, as an outline is created free of the demands of constructing a coherent text, one might be more likely to quickly jot down some ideas. Sentences in outlines also tend to be shorter, and hence there is less planning necessary for each sentence. Therefore, writing an outline might be easier than writing a text, which is why one does not need to pause as long between sentences compared to when no outline is made prior to writing.

Lastly, the results here showed that outlining significantly influenced the ratio long pauses between words. Compared to the non-outliners, the ratio of long pauses was significantly higher for the outliners during the first four time intervals of the writing process, whereas it was lower during the end. These findings are consistent with earlier work which showed that longer pauses are related to higher-level planning processes (Medimorec & Risko, 2017). These results could be interpreted as outlining being a factor that leads to more elaborate planning during the beginning of the writing process, which in turn, leads to less planning towards the end of it.

6.3 The effect of outlining on fluency and verbosity

With respect to the third sub question, the results indicated that outlining negatively affected the ratio characters typed in P-bursts during the outline phase. This study did not find significant differences between conditions during the time the subjects were working on their essay. Since longer production bursts have been associated with higher writing fluency (Van Waes & Leijten, 2015), these findings were against the expectations. These results might be related to the results of the ratio intext revisions feature. Since it was thought that outliners made more intext revisions due to a continuous reflection process, these intext revisions might cause the flow of writing to break more often. This could in turn lead to fewer p-bursts and p-bursts of shorter length.

The results of the last two features, the number of characters typed per minute and the number of keystrokes, were very similar. The multilevel models showed significant effects of outlining during the first two intervals, which implied that fluency is

considerably reduced when one is outlining. However, as indicated by the visualization, fluency was higher for the outline subjects when they were writing their essay. A comparison of these findings with those of other studies confirms that outlining does improve fluency during transcription time (Kellogg, 1988, 1990; Limpo & Alves, 2018), but does not improve fluency when measured over the total time (Johnson et al., 2012; Ong & Zhang, 2010; De Smet et al., 2014)).

To summarize, the current study shows the importance of analyzing the effects of outlining on the writing process over time, as temporal variations were found among almost all included keystroke features. Moreover, the results showed that outlining does not have the same effect on all aspects of writing. Revision and fluency features were mainly affected by outlining for the time intervals in which the outlining took place. By contrast, features related to pauses, in particular the mean time between sentences and the ratio long pauses between words, changed in almost all time intervals as a consequence of outlining. When taking all features together, the results suggest that outlining initially hampers the flow of writing but subsequently compensates this by making writing more efficient later on. This strengthens the findings of earlier work suggesting that outlining helps to separate the higher-level cognitive processes of planning and translating (Kellogg, 1990). Since the separation of these processes could reduce the amount of cognitive load experienced during writing (Galbraith & Rijlaarsdam, 1999; Kellogg, 1988, 1990), the results of this study suggest that in this way, outlining might be able to enhance the writing process.

6.4 Limitations and recommendations for future work

Several limitations were identified in the current study. First, the method used to split up the writing process into smaller sub phases is a point of discussion. In this study, the writing process was split up into evenly sized time intervals of 5 minutes. This was done as it allowed for easy interpretation of the results, with a clear distinction between the outlining phase and the writing phase. However, this method also has an important drawback since the results of participants who took different amounts of time to do the writing task were all analyzed together. This could mean for instance, that the fourth time interval represents the final stage of writing for participants that finished early, whereas it is somewhere in the middle of the writing process for participants that took the full time. As the writing behaviors of the participants might change depending on whether they are almost finished, or just halfway, this could have influenced the findings. A different option for defining time intervals used in other studies (Limpo & Alves, 2018; Van Waes & Leijten, 2015) is to split the writing process of each individual

up in a given number of intervals. The benefit of this approach is that one can easier generalize between writing phases. For example, if the writing process is split up into three phases, the third phase is for all participants the final phase. A disadvantage of this approach is that the results are harder to compare between participants, as participants that have a longer total time, will also have longer sub time intervals. A third approach seen in writing analytics (Conijn, Cook et al., 2020), is to split the process up in segments with an equal amount of keystrokes. The main benefit of this approach is that it solves problems when participants don't have data in certain time intervals. It also may lead to less extreme data points that could be caused by time intervals with only very little keystroke events. However, the results may also be harder to interpret as it could result in much variation in the time length of each interval. Since each approach has its pros and cons, it is important that these are carefully taken into account when interpreting the results. Future research could analyze how these different approaches can affect the outcomes of a study.

A second limitation of the current study is that participants could be assigned to a condition that is not in line with their personal writing preferences. That is, some participants might have been asked to create an outline prior to writing while not liking this. By contrast, participants in the non-outline condition might have made some sort of outline without explicitly being asked to do so. Such differences between individuals might have influenced part of the results. In the current study, these personal characteristics were considered by measuring the participants' preferred writing style using two scales developed by Kieft et al. (2008). However, as one scale was deemed unreliable and the other scale did not show any significant effects, the validity of these measures could be doubted. In an earlier study, De Smet et al. (2014) used the same questionnaires and found Cronbach's alpha values of only 0.63 and 0.71 for these scales, which also suggests that their reliability is not convincing. Therefore, future studies could develop more reliable measures to describe a persons preferred writing style. These can in turn be used in research, such as the current study, to control for potential confounding factors.

Lastly, a third limitation is that keystroke logging is a more indirect method to measure the writing process which relies on making inferences. An important reason for using keystroke logging in the current study was based on the temporal insights that it could provide. While it turned out that temporal analyses were useful for getting more detailed insights into how outlining works, linking the keystroke features directly to their underlying cognitive processes remains a challenge (Baaijen et al., 2012). To solve this, future work could combine keystroke logging with other, more direct, measurement methods such as self-report measures.

Even though not in the scope of the current study, future studies could also focus more on how the effects of outlining on the writing process are related to writing product quality. This study was the first to investigate outlining from a process perspective, and since the results are promising, also incorporating the writing product would make a logical next step. This way, one could study whether outlining has a moderating effect on product quality. Moreover, another interesting angle would be to take into the account the properties of the outlines. Since writing plans with higher levels of structure and a higher degree of detail are associated with higher essay scores (Chai, 2006), the quality of an outline could also have an impact on the writing process. Future studies that take such aspects into account might be able to create a more detailed view of how an outline benefits the writing process most.

7 Conclusion

The current study used a temporal analysis approach using keystroke logging to investigate the effects of outlining on the writing process. In particular, it analyzed how outlining affected measures of revision behavior, pausing behavior, and writing fluency. Despite not being completely decisive, the results of the multilevel models add to earlier literature suggesting that outlining might stimulate a separation of planning and translation processes during writing. By reducing conflict between cognitive processes, outlining has the potential to induce a general optimization of the writing process. Moreover, the significant findings emphasize the importance of investigating the writing process from a temporal perspective. Overall, this study adds to our understanding on the effect of outlining on the organization of the cognitive processes involved in writing. It paves the way for more studies on planning strategies, ultimately, hoping to support students in becoming better writers.

References

Allen, L. K., Jacovina, M. E., Dascula, M., Roscoe, R. D., Kent, K., Likens, A. D. & McNamara, D. S. (2016). ENTERing the time series SPACE: Uncovering the writing process through keystroke analyses. *Proceedings of the 9th International Conference on Educational Data Mining*, 22-29.

Baaijen, V. M. & Galbraith, D. (2018). Discovery through writing: Relationships with writing processes and text quality. *Cognition and Instruction*, *36*(3), 199–223. doi: 10.1080/07370008.2018.1456431

Baaijen, V. M., Galbraith, D. & de Glopper, K. (2012). Keystroke analysis: Reflections on procedures and measures. *Written Communication*, *29*(3), 246–277. doi: 10.1177/0741088312451108

Baaijen, V. M., Galbraith, D. & de Glopper, K. (2014). Effects of writing beliefs and planning on writing performance. *Learning and Instruction*, *33*, 81–91. doi: 10.1016/j.learninstruc.2014.04.001

Barkaoui, K. (2016). What and when second-language learners revise when responding to timed writing tasks on the computer: The roles of task type, second language proficiency, and keyboarding skills. *The Modern Language Journal*, *100*(1), 320–340. doi: 10.1111/modl.12316

Beauvais, C., Olive, T. & Passerault, J.-M. (2011). Why are some texts good and others not? relationship between text quality and management of the writing processes. *Journal of Educational Psychology*, *103*(2), 415–428. doi: 10.1037/a0022545

Braaksma, M. A., Rijlaarsdam, G., van den Bergh, H. & van Hout-Wolters, B. H. (2004). Observational learning and its effects on the orchestration of writing processes. *Cognition and Instruction*, *22*(1), 1–36. doi: 10.1207/s1532690xci2201_1

Broekkamp, H. & Van Den Bergh, H. (1996). Attention strategies in revising a foreign language text. In G. Rijlaarsdam, H. Van Den Bergh & M. Couzijn (Eds.), *Theories, models and methodology in writing research* (p. 170-181). Amsterdam: Amsterdam University Press.

Chai, C. (2006). Writing plan quality: Relevance to writing scores. *Assessing Writing*, *11*(3), 198–223. doi: 10.1016/j.asw.2007.01.001

Chukharev-Hudilainen, E., Saricaoglu, A., Torrance, M. & Feng, H.-H. (2019). Combined deployable keystroke logging and eyetracking for investigating l2 writing fluency. *Studies in Second Language Acquisition*, *41*(3), 583–604. doi: 10.1017/s027226311900007x

Coirier, P., Andriessen, J. E. B. & Chanquoy, L. (2000). From planning to translating: the specificity of argumentative writing. In P. Coirier & J. E. B. Andriessen (Eds.), *Foundations of argumentative text processing* (p. 1-29). Amsterdam, the Netherlands: Amsterdam University Press.

Conijn, R. (2020). *The keys to writing: A writing analytics approach to studying writing processes using keystroke logging.* Tilburg University.

Conijn, R., Cook, C., Van Zaanen, M. & Van Waes, L. (2020). Early prediction of writing quality using keystroke logging. *The Keys to Writing: A writing analytics approach to studying writing processes using keystroke logging*.

Conijn, R., Dux Speltz, E., van Zaanen, M., van Waes, L. & Chukharev-Hudilainen, E. (2020). A product and process oriented tagset for revisions in writing. *The Keys to Writing: A writing analytics approach to studying writing processes using keystroke logging*.

Conijn, R., Roeser, J. & van Zaanen, M. (2019). Understanding the keystroke log: The effect of writing task on keystroke features. *Reading and Writing*, *32*(9), 2353–2374. doi: 10.1007/s11145-019-09953-8

De La Paz, S. & Graham, S. (2002). Explicitly teaching strategies, skills, and knowledge: Writing instruction in middle school classrooms. *Journal of Educational Psychology*, *94*(4), 687–698. doi: 10.1037/0022-0663.94.4.687

De Smet, M. J. R., Brand-Gruwel, S., Broekkamp, H. & Kirschner, P. (2012). Write between the lines: Electronic outlining and the organization of text ideas. *Computers in Human Behavior*, *28*(6), 2107–2116. doi: 10.1016/j.chb.2012.06.015

De Smet, M. J. R., Brand-Gruwel, S., Leijten, M. & Kirschner, P. A. (2014). Electronic outlining as a writing strategy: Effects on students' writing products, mental effort and writing process. *Computers & Education*, *78*, 352–366. doi: 10.1016/j.compedu.2014.06.010

De Smet, M. J. R., Broekkamp, H., Brand-Gruwel, S. & Kirschner, P. (2011). Effects of electronic outlining on students' argumentative writing performance. *Journal of Computer Assisted Learning*, *27*(6), 557–574. doi: 10.1111/j.1365-2729.2011.00418.x

Ellis, R. & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, *26*(01). doi: 10.1017/s0272263104026130

Faul, F., Erdfeldner, E., Lang, A. G. & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*(39), 175-191.

Galbraith, D. (1992). Conditions for discovery through writing. *Instructional Science*, *21*, 45-72. doi: 10.1007/bf00119655

Galbraith, D., Ford, S., Walker, G. & Ford, J. (2005). The contribution of different components of working memory to knowledge transformation during writing. *L1-Educational Studies in Language and Literature*, *5*(2), 113–145. doi: 10.1007/s10674-005-0119-2

Galbraith, D. & Rijlaarsdam, G. (1999). Effective strategies for the teaching and learning of writing. *Learning and Instruction*, *9*(2), 93–108. doi: 10.1016/s0959-4752(98)00039-5

Galbraith, D. & Torrance, M. (2004). Revision in the context of different drafting strategies. *Studies in Writing*, 63–85. doi: 10.1007/978-94-007-1048-1_5

Graham, S., McKeown, D., Kiuhara, S. & Harris, K. R. (2012). Meta-analysis of writing instruction for students in elementary grades. *Journal of Educational Psychology*, *104*(4), 896–896. doi: 10.1037/a0029939

Graham, S. & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, *99*(3), 445–476. doi: 10.1037/0022-0663.99.3.445

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

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

Hayes, J. R. & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. Steinberg (Eds.), *Cognitive processes in writing* (p. 3-30). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Hayes, J. R. & Nash, J. G. (1996). On the nature of planning. In M. C. Levy & S. E. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (p. 29-55). Routledge, Taylor & Francis Group. New York.

Isnard, N. & Piolat, A. (1994). The effects of different types of planning on the writing of argumentative text. In G. Eigler & T. Jechle (Eds.), *Writing: Current trends in european research* (p. 121-132). Freiburg, Germany: Hoochschul Verlag GmbH.

Johnson, M. D., Mercado, L. & Acevedo, A. (2012). The effect of planning sub-processes on l2 writing fluency, grammatical complexity, and lexical complexity. *Journal of Second Language Writing*, *21*(3), 264–282. doi: 10.1016/j.jslw.2012.05.011

Kellogg, R. T. (1988). Attentional overload and writing performance: Effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(2), 355–365. doi: 10.1037/0278-7393.14.2.355

Kellogg, R. T. (1990). Effectiveness of prewriting strategies as a function of task demands. *The American Journal of Psychology*, *103*(3), 327. doi: 10.2307/1423213

Kellogg, R. T. (1993). Observations on the psychology of thinking and writing. *Composition Studies*, *21*(1), 3-41.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (p. 57-72). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kellogg, R. T. & Whiteford, A. P. (2009). Training advanced writing skills: The case for deliberate practice. *Educational Psychologist*, *44*(4), 250–266. doi: 10.1080/00461520903213600

Kieft, M., Rijlaarsdam, G. & van den Bergh, H. (2008). An aptitude–treatment interaction approach to writing-to-learn. *Learning and Instruction*, *18*(4), 379–390. doi: 10.1016/j.learninstruc.2007.07.004

Lea, M. R. & Street, B. V. (1998). Student writing in higher education: An academic literacies approach. *Studies in Higher Education*, *23*(2), 157–172. doi: 10.1080/03075079812331380364

Leijten, M. & Van Waes, L. (2013). Keystroke logging in writing research. *Written Communication*, *30*(3), 358–392. doi: 10.1177/0741088313491692

Leijten, M., Van Waes, L., Schriver, K. & Hayes, J. (2014). Writing in the workplace: Constructing documents using multiple digital sources. *Journal of Writing Research*, *5*(3), 285–337. doi: 10.17239/jowr-2014.05.03.3

Lillis, T. & Turner, J. (2001). Student writing in higher education: Contemporary confusion, traditional concerns. *Teaching in Higher Education*, *6*(1), 57–68. doi: 10.1080/13562510020029608

Limpo, T. & Alves, R. A. (2018). Effects of planning strategies on writing dynamics and final texts. *Acta Psychologica*, *188*, 97–109. doi: 10.1016/j.actpsy.2018.06.001

Lindgren, E., Westum, A., Outakoski, H. & Sullivan, K. P. (2019). Revising at the leading edge: Shaping ideas or clearing up noise. In (p. 346 - 365). Leiden, The Netherlands: Brill. doi: https://doi.org/10.1163/9789004392526_017

Medimorec, S. & Risko, E. F. (2017). Pauses in written composition: on the importance of where writers pause. *Reading and Writing*, *30*(6), 1267–1285. doi: 10.1007/s11145-017-9723-7

Olive, T. (2014). Toward a parallel and cascading model of the writing system: A review of research on writing processes coordination. *Journal of Writing Research*, *6*(2), 173–194. doi: 10.17239/jowr-2014.06.02.4

Olive, T. & Kellogg, R. T. (2002). Concurrent activation of high- and low-level production processes in written composition. *Memory & Cognition*, *30*(4), 594–600. doi: 10.3758/bf03194960

Olive, T., Kellogg, R. T. & Piolat, A. (2002). The triple task technique for studying the process of writing. *Studies in Writing*, 31–59. doi: 10.1007/978-94-010-0468-8_3

Ong, J. & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in efl students' argumentative writing. *Journal of Second Language Writing*, *19*(4), 218–233. doi: 10.1016/j.jslw.2010.10.003

Page-Voth, V. & Graham, S. (1999). Effects of goal setting and strategy use on the writing performance and self-efficacy of students with writing and learning problems. *Journal of Educational Psychology*, *91*(2), 230–240. doi: 10.1037/0022-0663.91.2.230

R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rogers, L. A. & Graham, S. (2008). A meta-analysis of single subject design writing intervention research. *Journal of Educational Psychology*, *100*(4), 879–906. doi: 10.1037/0022-0663.100.4.879

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

Strömqvist, S., Homqvist, K., Johansson, V., Karlsson, H. & Wengelin, A. (2006). What keystroke logging can reveal about writing. In K. Sullivan & E. Lindgren (Eds.), *Computer keystroke logging and writing: methods and applications* (p. 45-71). Elsevier.

Torrance, M. & Galbraith, D. (2006). The processing demands of writing. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (p. 67-80). The Guilford Press.

Torrance, M., Thomas, G. & Robinson, E. (1994). The writing strategies of graduate research students in the social sciences. *Higher Education*, *27*(3), 379–392. doi: 10.1007/bf03179901

Torrance, M., Thomas, R. & Robinson, E. (2000). Individual differences in undergraduate essay-writing strategies: A longitudinal study. *Higher Education*, *39*(2), 181-200.

Van Waes, L. & Leijten, M. (2015). Fluency in writing: A multidimensional perspective on writing fluency applied to l1 and l2. *Computers and Composition*, *38*, 79–95. doi: 10.1016/j.compcom.2015.09.012

Van Waes, L., van Weijen, D. & Leijten, M. (2014). Learning to write in an online writing center: The effect of learning styles on the writing process. *Computers & Education*, *73*, 60–71. doi: 10.1016/j.compedu.2013.12.009

Walvoord, B. E., Anderson, V. J., Breihan, J. R., McCarthy, L. P., Robison, S. M. & Sherman, A. K. (1995). Functions of outlining among college students in four disciplines. *Research in the Teaching of English*, *29*(4), 390-421. doi: 10.1002/ets2.12075

Appendix A - Power analysis

Prior to the data collection for this master thesis, a power analysis was performed to determine the amount of participants necessary to ensure the study had enough power. However, as the desired number of participants was not reached, the design of the study was adapted to that of the current study. Below are the properties and results of the original power analysis.

t tests – Means: Difference between two independent means (two groups)

Analysis:	A priori: Compute required sample size		
Input:	Tail(s)	=	Two
	Effect size d	=	0.8
	α err prob	=	0.05
	Power (1−β err prob)	=	0.9
	Allocation ratio N2/N1	=	1
Output:	Noncentrality parameter δ	=	3.2984845
	Critical t	=	1.9965644
	Df	=	66
	Sample size group 1	=	34
	Sample size group 2	=	34
	Total sample size	=	68
	Actual power	=	0.9015019

Appendix B - Writing task

The writing task was as follows:

Write an argumentative essay about the following topic: To bring the COVID-19 pandemic to an end, countries are busy vaccinating their inhabitants. Since getting a vaccine is not obligated, governments and businesses are increasingly looking for ways to tell who has been vaccinated and who is not. This pursuit has stirred up discussions about the so-called "vaccine passport", which is a verified proof to show that you received a vaccine. If event organizers are allowed to ask for vaccine passports, this could mean that people who have not been vaccinated could be denied entry. Therefore, the topic for you essay is as follows: **"People who cannot show a vaccine passport at the entry of events should not be allowed to access."**

What is your opinion? Use specific reasons and examples to support your answer.

Write at least 200 words.

Appendix C - Task instructions

Based on the instructions given in Kellogg (1990), the specific outling instructions were formulated as follows:

Outlining

As mentioned [in the general instructions], you have to use the planning technique of outlining. With outlining, you generate and organize ideas for your essay before you actually start with writing. For this task, you should create a standard hierarchical outline using multilevel bulletpoints in which you distinghuish between main ideas, subpoints, further subpoints and so on. This can be done as follows:

- Main idea
  - First subpoint
    - Further subpoints
  - Second subpoint
    - Further subpoints
- Etc.

Your outline may contain as many points and as many levels as you wish.

Appendix D - Measuring preferred writing style

Items from the scales developed by Kieft et al. (2008) to measure preferred writing style.

Revising items (15)

1. Before I hand in my text, I check whether it is structured logically
2. *I don't pay much attention to whether I'm satisfied with my text myself
3. Writing helps me to clarify my thoughts
4. When I write a text, I question myself from time to time whether the text is comprehensible for my readers
5. *When I write a text, I find it difficult to form ideas about which I can write
6. Before I start to write a text, I prefer to write down some thoughts on a scribbling paper to discover what I think about the topic
7. While writing, I regularly check whether my text doesn't contain sentences that are too long or incorrect
8. *I don't pay much attention to skipping sentences or thoughts
9. When I reread and rewrite my text, the structure of the text may change a lot
10. When I rewrite my texts, the content often changes a lot
11. *I usually hand in my text without checking whether the paragraphs are well arranged
12. When I have finished writing, I reread and improve a lot: this may change a lot in my text
13. I have to reread the texts I wrote, to prevent redundancies
14. *Usually, the texts I write are not very creative
15. When I know what to write globally, I write my texts very easily

Planning items (11)

1. *Planning a text is not useful for me
2. *When I start writing, I don't know what the content of the text will be
3. When I write a text, I spend a lot of time thinking on how to approach it
4. Before I start to write, it is clear for me what I want to achieve with my readers
5. Before writing a text, I jot down some notes on a scribbling paper. Later, I elaborate these notes
6. I always use a diagram before I start to write

7. Before I start to write, I have to know what the content of the text will be. Therefore, planning is important for my writing
8. I need to have my thoughts clear, before I can start to write
9. *When writing, I sometimes write paragraphs of which I know that they are not yet correct, but I prefer to continue writing
10. When I reread my texts, sometimes they are very chaotic
11. Before I write down a sentence, I have it clear in my mind

Items with a * must be recoded.

Appendix E - Questions regarding writing skills and experience

Self-report questions to measure writing skills and experience

1. I consider my academic writing skills to be (think of your writing skills for school assignments, reports etc.):
   (a) Very poor
   (b) Poor
   (c) Fair
   (d) Good
   (e) Excellent
2. Have you ever participated in extracurricular lessons to improve your writing skills (e.g. academic writing courses, writing workshops etc.)?
   (a) No
   (b) Yes, please specify:
3. How often do you write per month? Think of solid writings, text messages etc. don't count as writing.
   (a) Daily or almost daily
   (b) 2-3 times a week
   (c) Once a week
   (d) Less than once a week
   (e) Never
4. Do you have significant writing experiences from non-educational activities (e.g. writing for a blog, writing for a newspaper or magazine, having a job that involves much writing etc.)?
   (a) No
   (b) Yes, please specify: