THE EFFECTS OF TIME CONSTRAINTS, GENRE, AND PROFICIENCY ON L2 WRITING
FLUENCY BEHAVIORS AND LINGUISTIC OUTCOMES

By

Jongbong Lee

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Second Language Studies – Doctor of Philosophy

2019

ABSTRACT

THE EFFECTS OF TIME CONSTRAINTS, GENRE, AND PROFICIENCY ON L2 WRITING
FLUENCY BEHAVIORS AND LINGUISTIC OUTCOMES

By

Jongbong Lee

Length of writing has been measured to identify development, and task and genre effects in second language (L2) writing. Moving beyond a singular focus on assessing writing outcomes (i.e., the length of writing), this study investigates L2 learners' writing fluency-related behaviors and the cognitive processes behind them by exploring the effects of genre, time constraints, and proficiency. Drawing on Kellogg's model of writing (1996), this study adopts a mixed-methods design and uses (1) keystroke logging to capture writing behaviors, such as fluency, pausing, and revision, (2) syntactic complexity analyzer and Coh-metrix to investigate linguistic complexity, and (3) stimulated recalls to reveal cognitive processes used by L2 learners.

Participants included 123 English L2 learners studying at a university, with high-intermediate (60 participants) or advanced (63 participants) proficiency according to standardized tests and a cloze test. Their writing behaviors were recorded by Inputlog 7.0, a keystroke logging program. The participants were assigned at random to the long-timed (60 minutes) or short-timed group (30 minutes). Furthermore, each participant was randomly assigned to either the narrative or the argumentative essay on the first day, and the other genre on the second day. Sixteen participants were randomly selected for stimulated recall sessions, and they were required to recall their writing processes as prompted by the screen recordings. For triangulating the data, this study used the stimulated recall comments and the keystroke logs. Additionally, the participants completed an exit survey which captured their perception on genres and time allotment.

Repeated measures MANOVAs revealed that the L2 learners' writing behaviors such as fluency and linguistic outcomes were affected by differences in time constraints, genre, and proficiency. The time constraints affected writing fluency behaviors in that learners in the short-timed group produced higher writing fluency behaviors, such as longer P-burst length than those in the long-timed group. The argumentative genre led the participants to respond with more complex language and less fluent writing behaviors than the narrative genre. The advanced learners showed more syntactically complex language and more fluent writing behaviors than the high intermediate learners. The stimulated recall data showed that L2 learners' writing processes, such as planning and translation, differed across time constraints, genre, and proficiency.

In addition, a two-way ANOVA showed that the effect of proficiency on writing quality was significant whereas the different time constraints did not affect writing quality. Writing fluency measures were correlated with linguistic measures and writing quality. A linear regression analysis showed that some writing fluency behavior measures predicted writing quality. Further, depending on proficiency and time allotment, the participants' perception on writing tasks differed. Taken together, the findings regarding theoretical, methodological, and pedagogical implications are discussed.

ACKNOWLEDGEMENTS

I have been so fortunate to have had so much support from many people. I would like to express gratitude to everyone I met during my graduate studies.

First, I would like to express deep thanks to Dr. Charlene Polio, whose exemplary supervision has enabled me to enjoy PhD life. She always reads my papers with lighting speed and gives me many helpful comments. She is aware of my strengths and weaknesses, and her good advice has helped me to develop my research. She genuinely cares about her students and is an inspiring role model and scholar. I hope to follow in her footsteps and become a scholar devoted to research, teaching, and mentoring students.

I also thank Dr. Shawn Loewen for helping me expand my ideas about writing fluency and add an exit questionnaire to this dissertation. In addition to offering his support for my dissertation, he read my two qualifying papers and gave me constructive feedback during my doctoral studies. I thank Dr. Paula Winke for providing me with relevant literature for this dissertation and giving me the opportunity to develop the theoretical background of this dissertation in her Language Assessment class. I would like to thank Dr. Patti Spinner for equipping me with research skills through her Advanced Topic in SLA class, which provided a foundation for my dissertation.

Although he was not on my committee, Dr. Peter De Costa has been supportive throughout my doctoral studies, and I am grateful that I have had many opportunities to work with him. I am also thankful to Dr. Ok-Sook Park in the Korean program for giving me the opportunity to teach Korean.

I would like to thank the College of Arts and Letters and the Graduate College for a

son and brother. Last but not least, I thank my wife, Myeongeun, for reading my dissertation, giving me constructive feedback, and teaching me what true love is. Thank you for all that you have done to help me during this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

Fluency has been used as a measurement of second-language performance and second-language (L2) development. It has also been used to better understand how specific tasks or genres affect L2 performance. Although there are different definitions of fluency, the term is generally considered to describe the flow and smoothness of language production (Koponen & Riggenbach, 2000; Segalowitz, 2010). For instance, Lennon (1990) considered oral fluency to be a global ability and a temporal aspect of performance. Schmidt (1992) suggested that fluency in speech production is an automatic procedural skill that shows how well learners perform when doing a task in real time.

In the assessment of writing fluency, the definition is narrowed down to specific, measurable characteristics. For example, traditionally, writing fluency is measured by the number of words and structures produced within a limited time, which is equivalent to temporal measures for oral language (Wolfe-Quintero, Inagaki, & Kim, 1998). These measures, however, may not reflect learners' writing performance perfectly, in part because writing behaviors such as pausing and revision may affect learners' time producing language (Abdel Latif, 2013; Kellogg, 1996).

Accordingly, the traditional method of assessing writing fluency is controversial (e.g., Abdel Latif, 2013; Van Waes & Leijten, 2015). Abdel Latif (2013) questioned the validity of previous studies that used the traditional method of dividing the number of words written within a given time frame. One reason for this is that L2 learners may pause in different places and for different reasons when writing, so not all pauses should be considered equal. In short, as writing fluency is affected by different writing processes such as monitoring, a single measure of length

(i.e., the number of words) may not fully capture fluency (Abdel Latif, 2013).

In order to validly measure writing fluency, recent researchers have explored new methods such as keystroke logging (e.g., de Smet, Leijten, & Van Waes, 2018; Révész, Kourtali, & Mazgutova, 2017). Keystroke logging is a useful tool for examining specific aspects of L2 writing. Concurrent and unobtrusive, it can record, for example, the length and timing of pauses. However, keystroke logging reveals only some writing behaviors, and so it alone cannot fully capture L2 writers' processes or their internal cognition during writing. Therefore, other methods such as stimulated recall (e.g., Révész, Kourtali, & Mazgutova, 2017) and think alouds (e.g., Schrijver, Van Vaerenbergh, & Van Waes, 2012) should be used to complement keystroke logging, thus compensating for some of its shortcomings, especially its inability to take account of the writers' thought processes (Geisler & Slattery, 2007, p. 197).

One important issue in the measurement of fluency is related to whether or not the writing is produced in a timed setting because the construct of fluency is affected by the amount of writing time available. When time limits are used, the question of how to appropriately set them for various tasks arises. In addition, although timed writing is used in many instructional and testing settings, many scholars have suggested that writing under time pressure is unnatural (e.g., Cho, 2003; Weigle, 2002). For instance, Weigle (2002, p. 172) pointed out the limitations of the timed impromptu essays that are widely used in testing and research to indicate L2 learners' development or production. She suggested that alternatives to short timed essays should be considered and that untimed essay writing gives L2 learners less anxiety and allows them time to generate ideas and to prepare to write about specific topics. While most previous studies investigating L2 writing have employed short timed writing tasks (e.g., 30-minute essays), the use of untimed or longer timed tasks is ecologically valid, and research findings based on such

tasks could be extended to real-life and instructional settings (Polio & Friedman, 2017; Polio & Lee, in press). Both timed and untimed essays are considered meaningful tasks, but few studies have investigated differences in how the different time constraints affect L2 learners' fluency, writing behaviors, and linguistic outcomes.

Fluency has also been used as a dependent variable in research that investigates task and genre effects. Researchers have employed a variety of theoretical frameworks to assess constructs of writing such as fluency and complexity while exploring the effects of different tasks and genres on L2 writing. For instance, Robinson's (2001) cognition hypothesis and Skehan's (1996) trade-off hypothesis both suggest that the linguistic complexity, accuracy, and fluency (CAF) of L2 learners' production are influenced by tasks. In addition, some genre-based studies (e.g., Lu, 2011) have shown that genres affect learners' production in terms of CAF. Although these previous studies used different frameworks, all of them have emphasized fluency as one of the constructs that help researchers find out how different types of writing affect L2 learners' production.

The primary goal of this dissertation is to delve into the interplay between time constraints, genre, proficiency, and linguistic outcomes. It will also contribute to previous research by examining writing fluency with different measures. To date, most previous studies that have examined how different aspects of writing tasks such as genre and time constraints affect L2 writers' fluency have used only product-based measures (e.g., the total number of words produced in a given time). Only a few studies (e.g., Kellogg, 1990; Révész, Kourtali, & Mazgutova, 2017) have examined how different task types affect writing fluency behaviors and the underlying cognitive processes of writing; however, these studies have used only short-timed tasks and a single task type. To address this research gap, this study uses a range of diverse

writing fluency measures including process-based measures (e.g., P-bursts), and connects writing fluency behaviors to cognitive processes in two different types of writing. In addition, the study explores how different genres and time constraints affect the processes and products of L2 writers at different proficiency levels. The results of the study's investigation of L2 writing fluency behaviors should provide theoretical and methodological implications for L2 writing research as well as L2 writing pedagogy and assessment.

The remainder of this dissertation is organized as follows. Chapter 2 reviews the literature on the relationship of fluency to complexity and factors responsible for differences in writing fluency and linguistic outcomes to explain the theoretical background for the study. It also presents the study's research questions. Chapter 3 describes the study's methodology. Chapter 4 presents the results of the analysis of the data, and Chapter 5 discusses these results with regard to the research questions. Chapter 6 concludes the dissertation, pointing out limitations of this study and suggesting some directions for future research.

CHAPTER 2. LITERATURE REVIEW

2.1. Definitions of fluency and its relationship to other measures

Fluency is often discussed along with other constructs of production such as complexity and accuracy. Many researchers have investigated second language learners' production in terms of the three CAF constructs, and the three constructs in CAF are interwoven with each other (Foster & Skehan, 1996). The constructs have been used to measure distinct components of L2 performance that may be manifested by L2 learners under different task conditions (e.g., Housen & Kuiken, 2009; Housen, Kuiken, & Vedder, 2012).

An underlying assumption of the three constructs is that L2 learners show development in the target language over time. In other words, proficient L2 learners tend to show more complex, accurate, and fluent writing than novice L2 learners. Another assumption is that the three constructs of CAF are influenced by writing task types. In most second language acquisition (SLA) studies, in addition to being used as indices of L2 development, the three constructs have been utilized to look for effects of pedagogical treatments and genre differences. In the following sections, the constructs of fluency, complexity, and accuracy, and how they have been used in L2 research, will be further discussed.

2.1.1. *Fluency*

Since ways of measuring speaking fluency have influenced ways of measuring writing fluency, the latter are defined and operationalized in varied ways. For example, Wolfe-Quintero

et al. (1997) defined fluency as a number of words or structures included in writing within a limited time. On the other hand, Snellings, Van Gelderen, and De Glopper (2004) defined fluency as the speed of lexical retrieval in writing. Recently, in a study on process-based writing fluency, Van Waes and Leijten (2015) proposed a multidimensional fluency model. They argued that writing fluency includes production, process variation, revision, and pause behavior, and that these four components can distinguish fluent and less fluent writers. By using principal component analysis, they confirmed that the four components together contribute to the multidimensional fluency model. They suggested that various components of writing fluency be examined in experimental settings for comparison between groups or tasks. In short, fluency has been defined in many ways depending on multiple components within it, and the different definitions of writing fluency lead to the various measurements for assessing it.

Given that writing fluency does involve multiple components, operationalizations of writing fluency differ. The usual measures include counting the number of production units produced in a given time. According to Wolfe-Quintero et al. (1998), fluency should be measured by the number of words or structural units that a writer can produce in a particular period of time, rather than by the sophistication of the vocabulary or structures produced. In other words, more fluent writers can produce more words and structures in a given time whereas less fluent writers can produce fewer words and structures in a given time. The most widely used measure is the number of words divided by writing time (e.g., Sasaki & Hirose, 1996). Some studies include quantity of writing (Sasaki, 2004) and words per T-unit (Larsen-Freeman, 2006), but Norris and Ortega (2009) suggested that words per T-unit (i.e., words per a main clause plus any clauses dependent on it) should be considered a complexity measure. As this brief summary of the research suggests, there remains some confusion regarding how best to measure writing

fluency.

Abdel Latif (2013) pointed out the definitional confusion over writing fluency due to its multiple components, and raised a concern about product-based writing fluency assessment; that is, the practice of measuring fluency quantitatively in a finished product. In most previous studies, the researchers have counted words or calculated sentence length (Johnson, Mercado, & Acevedo, 2012) or composition rate (Sasaki, 2000). Few researchers have examined process-based measures such as pausing or length of translating episodes. To assess L2 learners' pausing, computer-based methods such as keystroke logging can be used, helping researchers assess L2 learners' real-time writing fluency (e.g., Leijten & Van Waes, 2006; Révész, Kourtali, & Mazgutova, 2017; Révész, Michel, & Lee, 2017, in press; Spelman Miller, Lindgren, & Sullivan, 2008; Van Hell, Verhoeven, & Van Beijsterveldt, 2008; Van Waes & Leijten, 2015). The computer-based methods have been used in both L1 and L2 writing research. The research that included the computer-based methods was involved in different research foci and languages (see Table 1). For instance, the software Inputlog (http://www.inputlog.net/) tracks writing activities by recording pauses, keystrokes, mouse action, and so on. The software can also calculate P-bursts, which are the units of text produced between pauses; that is, the number of typed characters between pauses. More fluent writers have fewer, longer P-bursts than less fluent writers (Chenoweth & Hayes, 2001; Van Waes & Leijten, 2015). These studies argue that both process-based and product-based measures should be considered in order to assess writing fluency accurately.

Table 1

Writing-Process Research Using Keystroke-Logging Techniques and Grouped by Research

Focus

Research focus	Language	Study
Writing fluency behaviors (e.g., pausing and revision)	L1 Portuguese	Alves et al. 2008
	L1 Dutch	Baaijen et al. 2012
		de Smet et al. 2018
		Leijten & Van Waes 2013
		Schrijver et al. 2012
	L1 Swedish	Wengelin et al. 2009
	L1 Russian	Chukharev-Hudilainen 2014
	L2 English	Lindgren & Sullivan 2003
		Spelman Miller 2005
		Révész, Michel, & Lee 2017
		Ranalli et al. 2018
		Ranalli et al. 2019
		Sullivan & Lindgren 2002
	L2 Swedish	Kowal 2014
	L2 French	New 1999
		Scott & New 1999
Task type comparison (e.g., genre)	L1 English	Chenoweth & Hayes 2003
		Eklundh & Kollberg 2003
		Eklundh 1994
		Deane et al. 2018
		Medimorec & Risko 2016
		Medimorec & Risko 2017
	L1 Dutch	de Smet et al. 2014
		Leijten, Van Waes, & Ransdell 2010
		Quinlan et al. 2012
		Van Waes & Schellens 2003
		Van Waes et al. 2010
	L1 German	Wallot & Grabowski 2013
	L2 English	Barkaoui 2015, 2016*
		Khuder & Harwood 2015
		Révész, Kourtali, & Mazgutova 2017
	L2 German	Thorson 2000*

Table 1 (cont'd)

Proficiency	L1 English and L2 German	Thorson 2000*
	L1 Dutch and L2 English	Stevenson et al. 2006
	L1 Dutch and foreign languages	Van Waes & Leijten 2015
	L1 and L2 English	Spelman Miller 2000
	L2 English	Barkaoui 2015, 2016*
		Ganem-Gutierrez & Gilmore 2018*
		Spelman Miller et al. 2008*
		Xu 2018
		Xu & Ding 2014
Writing quality	L1 English	Almond et al. 2012
		Deane 2014
		Zhang & Deane 2015
		Guo et al. 2018
	L2 English	Ganem-Gutierrez & Gilmore 2018*
		Spelman Miller et al. 2008*
		Révész, Michel, & Lee 2017*
Learning style	L1 Dutch	Van Waes, Van Weijen, & Leijten 2014

Note. * indicates that the study falls in more than one category.

To understand and assess writing fluency better, it is worthwhile to compare writing fluency with speaking fluency. One of the differences between writing and speaking fluency is related to processing (Abdel Latif, 2013). L2 learners' production behaviors are different in tasks that are the same except for modality. Speaking is generally faster than writing, and L2 speech can be analyzed by the temporal fluency measures of pausing and speech rate because it needs to be produced in a given time. On the other hand, L2 learners' fluency behaviors vary more in writing than in speaking; for example, some learners pause a lot when they are beginning to write, and then speed up, while others might do the opposite. These behaviors can be strategic or inconsistent, and pausing may support or hinder writing. Therefore, pausing while writing may not be a sign of dysfluency, unlike pausing while speaking. In addition, pausing at different

locations is often associated with planning or other writing processes (Schilperoord, 1996). According to Abdel Latif (2013), a valid measurement of writing fluency should take account of chunks or spans of text produced; that is, the "bursts" occurring between pauses (i.e., P-bursts). As mentioned above, writing fluency can be defined in different ways, which results in different measurements. Hence, including different measurements increases the validity of assessments of writing fluency. In addition, with the help of keystroke logging software, it is possible to measure how writers write and revise by examining the ratio of process and product.

Fluency can also be used to show L2 development over time (e.g., Spelman Miller et al., 2008; Yoon & Polio, 2017). For example, by examining the number of words produced, Yoon and Polio (2017) did not find significant differences between genres but they did find a difference over time. As with many other studies (e.g., Knoch, Rouhsahd, & Storch, 2014; Godfrey, Treacy, & Tarone, 2014; Knoch, Roushad, Oon, & Storch, 2015), the L2 learners in their study showed a significant increase in fluency over the course of one semester but notably did not improve in terms of accuracy. Spelman Miller et al. (2008) investigated writing fluency in a longitudinal study in terms of bursts (typed characters between pauses and/or revisions), and they measured fluency during bursts (writing time between pauses and/or revisions). Although theirs was a small-scale study, they showed that fluency and the length of writing bursts both increased over time.

2.1.1.1. Fluency and writing processes

Although both speaking and writing modes require productive skills, they differ crucially in processing time. Pausing and speech rate are key temporal elements in speaking, and they

affect the product's comprehensibility, whereas in writing, pausing and writing rate vary depending on a variety of factors, and are not directly visible in the final product. As mentioned briefly above, writing cannot be accurately assessed by product-based measures and pausing alone; such measures do not tell us much about differences in how shorter or longer texts are produced depending on tasks or learner factors (Abdel Latif, 2013). In contrast, process-based measures such as P-bursts, as recorded by keystroke logging, can capture more information about the cognitive processes that L2 learners engage in while performing writing. Therefore, by employing a varied array of measures, it is possible to more accurately examine the construct of fluency in the writing mode.

Nevertheless, keystroke logging cannot reveal L2 learners' internal cognitive processes. Although it allows a glimpse of where and what learners write quickly and slowly, and how they revise and pause, it does not explain why they do so. Recently, Révész, Kourtali, and Mazgutova (2017) tried to triangulate keystroke logging with other methods. They conducted stimulated-recall sessions with four students to find out where they paused and revised. Their participants were advanced proficiency L2 English users, and the authors used English for the stimulated-recall sessions, but it is worth considering whether the L1 might be more useful to elicit rich data (Gass & Mackey, 2017).

Several researchers have sought to identify and explain how cognitive processes are involved in writing processes (Flower & Hayes, 1981; Kellogg, 1996; Sasaki, 2000, 2004; Sasaki & Hirose, 1996). Assuming that writing is a complex process, Flower and Hayes (1981) and Kellogg (1996) proposed models of writing. Flower and Hayes broke down the writing process into nonlinear, interactive processes of planning, translating, and revising. For instance, when reading a passage one has written, one may notice and repair errors, or make changes while

planning the next step. Thus, writers can demonstrate pausing, deletion, insertion, and movement behaviors (Spelman Miller et al., 2008). The Flower and Hayes model considers task environment, cognitive processes involved in writing, and the writer's long-term memory. The task environment includes external factors that influence writing tasks such as time constraints. The cognitive processes in writing involve planning, translating, and revision. The long-term memory stores knowledge of the genre, of the topic, and of the audience. Kellogg's model also involves three processes, which he called formulating, executing, and monitoring. These labels suggest the interactive relationship between cognitive processes and linguistic encoding processes. Execution involves motoric skills such as handwriting and typing. Monitoring is done to check if the intended meaning has been delivered well. Formulation deals with planning ideas and translating them into linguistic expressions. Translating ideas into linguistic expressions includes subprocesses such as selecting lexical units and encoding syntactic structures. Kellogg suggested that the three processes are active simultaneously, and that the extent to which the three processes are achievable depends on learners' working memory. More specifically, the central executive in working memory is responsible for the processes of formulating and monitoring, but not executing. This writing model also predicts advantages for both text quality and fluency when writing tasks place fewer demands on working memory (e.g., by including extra planning/outlining time), because the quality and fluency of writing depends on formulation and monitoring processes (Kellogg, 1990).

These writing models do not explicitly relate task types to writing processes and production, but it is likely that L2 writing behaviors and fluency can be influenced by different genres or tasks (e.g., Hayes, 1996). When genres are not familiar or tasks are cognitively complex, it is possible that L2 learners may have difficulties due to limited working memory

(Kellogg, 1990, 1996; Révész, Kourtali, & Mazgutova, 2017). L2 learners may also feel pressured by limited writing time, which could force them to generate ideas from long-term memory. Such pressures can affect underlying cognitive processes such as translating and planning, resulting in slower processing. And slow processing in turn can lead to more pauses and revisions.

With respect to the relationship between these writing processes and writing behaviors such as pausing and revising, some previous studies have explored alternative research methods for delving into L2 fluency (e.g., Lindgren & Sullivan, 2006; Stevenson, Schoonen, & Glopper, 2006). For instance, Thorson (2000) utilized keystroke logging to compare participants' revision behaviors when writing in their L1 and in their L2, as well as when responding to two different genres. Recently, Révész, Kourtali, and Mazgutova (2017) adopted a process-oriented perspective on fluency to look for task effects. The study used pausing behaviors, total writing time divided by total number of words/characters excluding pauses (minutes per word and characters per word), the number of words/characters occurring between pauses (words per P-burst and characters per P-burst), and revision behaviors. They did not find task effects in terms of overall fluency but did find a task effect on pausing between sentences as well as on revision behaviors. They suggested that a more complex task (i.e., a task in which content was not provided) led to more extensive pausing at higher level discourse units such as sentences, and to more revisions below the word level. In addition to quantitative data, they collected qualitative data through stimulated-recall sessions to attempt to explain L2 learners' cognitive writing processes. By including both traditional fluency measures and process-based measures, they were able to shed light on task effects that might not be captured with traditional measures of fluency alone.

L2 learners' linguistic encoding processes also differ depending on their proficiency or development (e.g., Chenoweth & Hayes, 2001; Housen & Kuiken, 2009; Housen et al., 2012; Roca de Larios, Manchón, Murphy, & Marín, 2008; Wolfe-Quintero et al., 1997). In considering how and why they differ, researchers generally assume that L2 learners can write more fluently as they learn more of the target language and, therefore, more proficient learners are more fluent in their writing than less proficient learners. Nevertheless, it is also possible that more proficient writers with a reflective writing style can make longer pauses and look back more than less proficient writers while producing high-quality writing (Bereiter & Scardamalia, 2009). However, many writing tasks take place under time pressure or in a testing environment. This matters because, generally, proficiency affects the speed with which learners can retrieve language; therefore, some learners can write more in the same amount of time than other learners. An improved automatized process of retrieving language is one aspect of improved proficiency.

2.1.1.2. Fluency and writing quality

Previous studies have provided evidence of the relationship between writing quality and writing behaviors including fluency (e.g., Barkaoui & Knouzi, 2018; Ganem-Gutierrez & Gilmore 2018; Porte, 1996; Révész, Kourtali, & Mazgutova, 2017; Spelman Miller et al., 2008; Stevenson et al., 2006). For instance, Stevenson et al. (2006) explored how Dutch high school students' writing behaviors were related to the quality of the texts they produced. The students wrote four argumentative essays (two in their L1 and two in their L2 English) on computers as they did think-aloud. Four raters rated the essays on only two criteria: content and language use. The findings showed some relationship between text length and text quality, but no relationship

between writing quality and revision types, although the authors hypothesized that a type of low-level revision (i.e., at the word and clause level) may be related to writing quality. Although their study was important in showing the relationship between writing behaviors and writing quality, their use of scores on only content and language use may have affected their findings. In addition, Bowles (2010) suggested that thinking aloud during writing activities may hinder learners' writing process, although Godfroid and Spino's (2015) L2 reading research showed that thinking aloud may not be as problematic as Bowles indicated it would be.

Spelman Miller et al.'s (2008) study examined a variety of factors in Swedish high school learners' L2 writing quality. As mentioned above, Spelman Miller et al. showed that two fluency measures (bursts and fluency during bursts) strongly predicted text quality. However, they found no relationship between revision or pausing behaviors and text quality. Although their longitudinal study was insightful regarding L2 writing fluency, more research on this topic is worthwhile to gain a clearer understanding of what fluency measures are related to text quality.

2.1.2. *Complexity*

As one of the CAF measures, fluency is related to complexity (Norris & Ortega, 2009). Oh (2006), for example, offered relevant empirical evidence for the relationship between complexity and fluency. She found that two fluency measures—namely, the number of T-units and the number of clauses—were positively correlated with complexity measures—namely, the number of words per T-unit and the number of words per clause, respectively (see also Qin & Uccelli, 2016). In other words, development of the L2 learners' complexity leads to development in their fluency and vice versa. Given the mutual impacts of complexity and fluency on changes

in each other, L2 learners' writing fluency should be explored together with complexity and its effects.

According to Housen and Kuiken (2009), complexity usually refers to both task complexity and L2 complexity. L2 complexity can be divided into linguistic complexity and cognitive complexity (see Figure 1). Cognitive complexity may contribute to L2 learners' attention or perception of difficulty; it is the subjective difficulty of processing language when L2 learners perform language tasks. Assessments of linguistic complexity tend to try to tap into L2 learners' interlanguage system, which is commonly measured by the length, sophistication, and diversity of the language the learners produce. Researchers examine learners' L2 complexity to try to understand how it is influenced by tasks or how it develops over time.

```
                    ┌──────────────┐
                    │  Complexity  │
                    └──────────────┘
                     /            \
        ┌────────────────┐   ┌──────────────┐
        │ Task complexity│   │ L2 complexity│
        └────────────────┘   └──────────────┘
                               /          \
              ┌────────────────────┐  ┌────────────────────┐
              │Cognitive complexity│  │Linguistic complexity│
              └────────────────────┘  └────────────────────┘
```
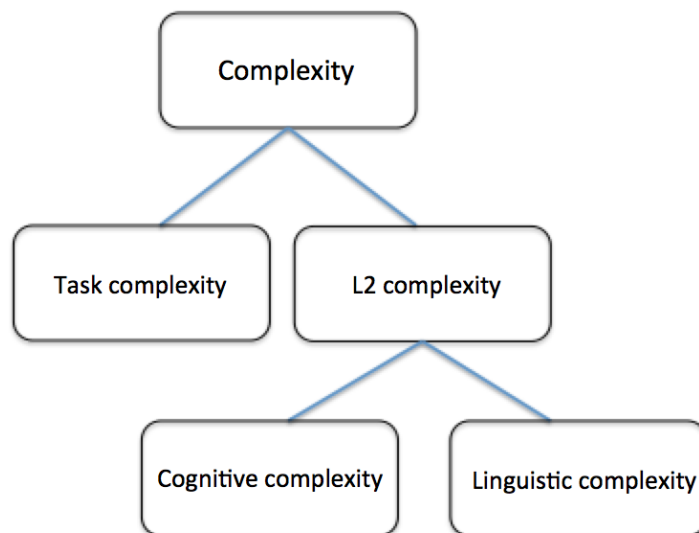
Figure 1. Complexity (Housen & Kuiken, 2009)

Previous studies have considered linguistic complexity in terms of syntactic complexity and lexical complexity (e.g., De Clercq & Housen, 2017; Housen, De Clercq, Kuiken, & Vedder, 2019; Norris & Ortega, 2009; Ortega, 2003). According to Norris and Ortega (2009), syntactic

complexity measures are often based on length, and calculated by dividing words by a chosen

production unit such as the sentence. They suggested that syntactic complexity should be

measured multidimensionally because L2 development cannot be explained by any single

measure, and the construct of syntactic complexity is composed of several subconstructs. In

other words, one syntactic complexity measure may not be enough to assess L2 learners'

development. For instance, Lu (2010, 2011) used 14 syntactic complexity measures to find genre

and proficiency differences in his automated text analysis; the different measures shed light on

the various characteristics of genre and proficiency (see Table 2). In a recent study, Kyle and

Crossley (2017) compared students' syntactic complexity and their verb argument construction

to the quality of their essays. They found that both types of index were significant predictors of

writing quality, although verb argument construction indices can explain a larger portion of

variance in writing quality than can syntactic complexity indices.

Lexical complexity can often be understood as lexical diversity, although there are many

other constructs (Norris & Ortega, 2009; Pallotti, 2015). A written text containing more different

vocabulary items can be deemed more complex than one with fewer. Several lexical complexity

measures exist, and there is some debate over which are best (McCarthy & Jarvis, 2010). For

example, the vocd-D index, a lexical diversity measure, has been considered a useful measure

that is not affected by text length as it is based on a mathematically probabilistic model (Malvern,

Richards, Chipere, & Durán, 2004), but McCarthy and Jarvis (2010) found that, in fact, it is

swayed by text length. Because of such uncertainty, analyses should include several different

measures of lexical complexity. Using a range of syntactic complexity and lexical complexity

measures makes it possible to investigate writing multidimensionally and may provide a clearer

analysis.

Table 2

Syntactic Complexity Measures (Lu, 2010)

	Measures	Definition
Type 1: Length of production unit	Mean length of sentence (MLS)	Number of words / Number of sentences
	Mean length of T-unit (MLT)	Number of words / Number of T-units
	Mean length of clause (MLC)	Number of words / Number of clauses
Type 2: Sentence complexity	Clauses per sentence (C/S)	Number of clauses / Number of sentences
Type 3: Subordination	T-unit complexity ratio (C/T)	Number of clauses / Number of T-units
	Complex T-unit ratio (CT/T)	Number of complex T-units / Number of T-units
	Dependent clause ratio (DC/C)	Number of dependent clauses / Number of clauses
	Dependent clauses per T-unit (DC/T)	Number of dependent clauses / Number of T-units
Type 4: Coordination	Coordinate phrases per clause (CP/C)	Number of coordinate phrases / Number of clauses
	Coordinate phrases per T-unit (CP/T)	Number of coordinate phrases / Number of T-units
	Sentence coordination ratio (T/S)	Number of T-units / Number of sentences
Type 5: Particular structures	Complex nominals per clause (CN/C)	Number of complex nominals / Number of clauses
	Complex nominals per T-unit (CN/T)	Number of complex nominals / Number of T-units
	Verb phrases per T-unit (VP/T)	Number of verb phrases / Number of T-units

The syntactic complexity and lexical complexity measures have been used to find task or genre differences because they show how well L2 writers deal with complex grammatical structures (e.g., Ellis & Yuan, 2004; Qin & Uccelli, 2016; Révész, Kourtali, & Mazgutova, 2017; Yoon & Polio, 2017). Yet there are mixed findings on the relationship between task complexity and linguistic complexity measures. Some researchers have found a positive relationship between them, but others have not. For example, Ellis and Yuan (2004) had three planning task conditions (pre-task planning, online planning, and no planning) in their experiment and found

that the L2 writers in the no planning condition (the most complex) produced less complex, accurate, and fluent writing than those in the other two conditions. They suggested that more complex tasks could elicit less complex language from L2 learners because the learners in no planning needed to formulate, execute and monitor their language under time pressure. Tavakoli (2014), on the other hand, found that storyline complexity did not affect written syntactic complexity. However, generally, in terms of genre, argumentative essays can elicit more complex language than narrative or descriptive essays (Biber & Conrad, 2009). The reason is that the communicative goals of argumentative essays require more complex structures and language than other genres of writing. Yoon and Polio (2017) found a strong genre difference in linguistic complexity, and suggested that argumentative essays can induce L2 learners as well as native speakers to produce more complex language than narrative essays. Yoon and Polio's comparisons between L2 learners' and native speakers' writing led them to suggest that the more complex language in argumentative essays can be attributed to the communicative functions of the genre rather than the possible reasoning demands of the genre. As these studies show, task and genre differences may be detected by measuring linguistic complexity, including syntactic and lexical complexity.

Previous studies have also used complexity measures to assess learners' development (e.g., Alexopoulou, Michel, Murakami, & Detmar, 2017; Beers & Nagy, 2011). For instance, Lu (2011) compared syntactic complexity in writing across four grade levels within the same institutions. He found that the first two adjacent levels (levels 1 and 2) and two or three pairs of nonadjacent levels could be distinguished by three length of production measures: mean length of clause, mean length of sentence, and mean length of T-unit. Unfortunately, he considered school level as equivalent to proficiency; his analysis could have been clearer if he had

administered a proficiency test such as the Test of English as a Foreign Language internet-based test (TOEFL iBT; www.ets.org). In addition, a longitudinal study is needed to capture learners' L2 development in terms of linguistic complexity.

2.2. Factors responsible for differences in writing fluency and linguistic outcomes: Time constraints, genre, and proficiency

Previous research shows that writing fluency is influenced by many variables such as proficiency, types of task, and writing conditions (e.g., Révész, Kourtali, & Mazgutova, 2017). For this reason, fluency can only be assessed fully by considering a variety of factors such as writing topics, genres, and writing time. Investigating the relationships among the variables is also essential in order to provide empirical evidence to see how genres, time constraints, and proficiency play a role in writing fluency. In this section, the relationships between fluency and time constraints, genre, and proficiency will be discussed.

2.2.1. *Time constraints*

According to Kellogg (1996), the time pressure to write rapidly can limit the central executive in terms of writing memory. Thus, increased time pressure inhibits smooth and responsive writing behavior; consequently, the writer may end up prioritizing formulation (i.e., planning and translating) over execution and monitoring. In other words, the amount of allowed time for a task can make a difference in the extent to which L2 learners stay at the formulation stage. This, in turn, may result in different lengths of pauses during writing, and consequently,

different writing fluency behaviors. In this regard, L2 learners' writing fluency should be investigated while taking into consideration time constraints.

In developing writing assignments or writing tests for L2 learners, time allocation is an issue in terms of outcome and process, including fluency (Caudery, 1990; Cho, 2003; Elder, Knoch, & Zhang, 2009; Knoch & Elder, 2010; Kroll, 1990; Lu, 2011; Polio & Glew, 1996; Powers & Fowles, 1996; Weigle, 2002). In her review article, Weigle (2002, p. 63) divided the dimension of time allowance into three sets (less than 30 minutes, 30–59 minutes, and 60–120 minutes). Wu and Erlam (2016) operationalized their study's timed condition by allowing 70% of the time the learner used on the untimed condition to examine the effect of time constraints on complexity, accuracy, fluency, and quality. Their findings showed that the learners produced more words in the untimed than in the timed condition. However, Elder, Knoch, and Zhang (2009) compared 30-minutes (short-timed) and 55-minutes (long-timed) writing tasks and did not find significant differences in terms of fluency ratings between them. In short, due to the differential operationalization of time constraints, previous studies have reported inconsistent findings across different time conditions in terms of L2 learners' performance.

In addition to comparing fluency, examining different linguistic features in L2 writing taps into other aspects of time-constraint effects. Some learners may benefit more from one or the other condition than other learners do. Younkin (1986), for example, compared native and nonnative English speakers' essays written in three different time conditions (no extra time, 10 minutes extra, and 20 minutes extra). He found that both the native and nonnative English groups benefited from the two extra time conditions. However, the essay test was part of a larger test, and thus it was hard to know how much time individual learners used for the essays. Ädel (2008) compared timed and untimed essays in corpora and argued that time can influence the proportion

of certain linguistic features such as first person singular pronouns. In a testing setting, Hale (1992) compared a test of written English in 30-minute and 45-minute conditions, and suggested that time allocation did not change performance on various test constructs although scores were higher in the longer condition. Powers and Fowles (1996) also compared graduate students' GRE writing in 40-minute and 60-minute conditions. Although the graduate students preferred and received a better score on untimed essays, the scores were not related to time allocation because scores under both conditions correlated similarly to nontest indicators of writing ability such as the students' reported success in various writing activities in college classes. Using corpus data, Lu (2011) compared timed and untimed argumentative essays in terms of seven syntactic complexity measures. The untimed essays elicited more syntactically complex language than the timed essays; however, Lu did not report how the corpus data he used operationalized timing conditions. More recently, Knoch and Elder (2010) found that test takers' scores were similar in two time conditions (55 minutes and 30 minutes), but they suggested that high proficiency learners benefited more from the extended time condition than did low proficiency learners, though they did not find significant differences in terms of quality. In short, the operationalization of time constraint conditions is different across the previous studies, and the effects of time constraint conditions remain inconclusive.

Only a few studies have delved into how different time constraints affect writers' production in relation to genres or tasks. For instance, Caudery (1990) compared two topics in time-restricted (timed, 40 minutes) and no-time-restricted (untimed, 1 hour) conditions. He did not find significant score differences between timed and untimed conditions. However, he compared only 12 students and did not report their proficiency levels. In addition, he claimed that eight of the students had written more slowly in the untimed condition, but he did not

measure their writing fluency accurately. More research is crucial to better understand whether and how the interplay of genres and time constraints affects L2 learners' writing, particularly their writing fluency or writing process.

2.2.2. *Genre*

Fluency is an important construct for understanding the effects of different genres on L2 learners' writing (e.g., Yang, 2014; Yang, Lu, & Weigle, 2015). Several studies have demonstrated that L2 learners' writing production can vary depending on genre (e.g., Jeong, 2017; Lu, 2011; Qin & Uccelli, 2016). With regard to writing processes, as the skill to deal with a certain genre increases, the effort needed to collect, plan, translate, and review decreases (Kellogg, 1994, p. 64). During writing, planning ideas, linguistically translating ideas or generating sentences, and reviewing ideas and text are all effortful; however, the pattern of differences among these processes varies with the task (Kellogg, 2001). Based on the assumptions of the writing models discussed in Section 2.2 (Flower & Hayes, 1980; Kellogg, 1996), previous researchers have found genre effects on L1 language processing and production, as shown by measures such as pause length (Beauvais, Olive, & Passerault, 2011; Medimorec & Risko, 2017; Van Hell et al., 2008) and text length (Beers & Nagy, 2011). In second language pedagogy and research, genre is also important for L2 writing theory and assessment regarding whether different genres elicit different processes and productions, such as more or less fluency from L2 learners.

Although many researchers have explored the effects of topic on L2 writing, only a few have specifically investigated the effects of genre on L2 learners' fluency (e.g., Ruiz-Funes,

2014, 2015; Thorson, 2000; Yang, 2014; Yoon & Polio, 2017). For example, Yoon and Polio (2017) examined fluency in narratives and argumentative essays as measured by total number of words produced in 30 minutes, but did not find a significant difference between the two genres, although the ESL learners in their study produced more words in narratives than in argumentative essays. Way, Joiner, and Seaman (2000) compared L2 French learners' 30-minute writing in three different genres. They found that the learners' narrative essays and expository essays were shorter than their descriptive essays.

In addition to its effects on fluency, genre plays an important role in aspects of L2 learners' written production (e.g., Lu, 2011; Qin & Uccelli, 2016; Way et al., 2000; Yoon, 2017). Examining different measures in addition to fluency is necessary to explain the multiple aspects of genre effects. As briefly mentioned above, Yoon and Polio (2017) found increased linguistic complexity (length of unit, coordination, particular structures, and lexical complexity) in argumentative essays, when compared to narratives; however, they did not find significant genre effect on fluency. On the other hand, Qin and Uccelli (2016) found more complexity and fluency in Chinese EFL learners' argumentative essays than in narratives in terms of number of words, lexical complexity, and number of words per clause. They also examined whether linguistic complexity features and fluency in argumentative essays and narratives were related to writing quality. The authors found that lexical complexity, syntactic complexity, and fluency were correlated to the quality of the argumentative essays and narratives. For both the genres, text length was found to be a strong predictor of quality. Although their use of holistic scores for writing quality allowed them to offer only a limited explanation of the relationship between the fluency measure and quality, their findings suggested that the L2 learners seemed to use different linguistic and discourse features to meet each genre's communicative purposes. Based on this

empirical evidence of L2 learners' use of complex and fluent language and the relationship

between fluency measure and writing quality in the two genres, these researchers suggested that

L2 learners use linguistic resources differently to fulfill different communicative purposes and

the functions of different genres (Biber & Conrad 2009; Biber, Gray, & Poonpon, 2011).

2.2.3. *Proficiency*

Fluency is used as an indicator of L2 proficiency as well as L2 development (e.g.,

Chenoweth & Hayes, 2001; Lambert & Kormos, 2014; Larsen-Freeman, 2006). Foreign

language fluency is connected to general proficiency and metalinguistic knowledge (Kowal,

2014; Wolfe-Quintero et al., 1998). As learners' proficiency develops, they gain greater ability to

monitor their language and pay attention to form while writing. In other words, L2 learners'

writing is affected by the amount of attention they have available for higher level processing

such as planning, generating ideas, or organizing content (Chenoweth & Hayes, 2001; Dekeyser,

2005).

Researchers have explored the relationship between proficiency and fluency. For instance,

Sasaki (2004) examined 11 participants' proficiency over three and a half years (including study

abroad experience) and found improvement in their fluency as measured by mean total number

of words and mean number of words per minute in their production. Taking a case study

approach, Thorson (2000) compared L1 and L2 essays and two different genres of writing

(articles and letters). The participants revised proportionally more when they wrote in L2

German than when they wrote in their L1 English. However, no clear genre effects were found in

their revision behaviors. Way et al. (2000) compared French level 1 and level 2 students' writing

and found that level 2 learners wrote more fluently than level 1 learners when fluency was measured by the number of words produced. In language testing, Barkaoui (2016) compared low proficiency and high proficiency learners' revision behaviors related to fluency. The study found that low proficiency learners made significantly more revisions than high proficiency learners because the high proficiency learners did not need to revise as often even though they wrote more than the low proficiency learners. Van Waes and Leijten (2015) examined participants' L1 (Dutch) and L2 (English, French, Spanish, or German) expository essay writing in terms of product-based and process-based fluency. They found that writing fluency significantly differed between the L1 and the L2. For example, the participants needed less pausing time between words when they wrote in the L1 than when they wrote in the L2. Although the participants' L2s were different, and the study does not attempt to explain the differences in pausing behavior, the study demonstrated that L1 writing fluency and L2 writing fluency differ in terms of different fluency measures.

Previous studies suggest that proficiency and genre together play an important role in L2 learners' linguistic outcomes, including fluency (e.g., Jeong, 2017; Qin & Uccelli, 2016; Ruiz-Funes, 2014, 2015). For instance, Ruiz-Funes (2015) examined intermediate and advanced learners' essays and found an interaction effect between genre and proficiency. Advanced learners wrote argumentative essays and contrast-compare essays while intermediate learners wrote expository essays and narratives. Ruiz-Funes suggested that argumentative and expository essays were more difficult than the other genres for each proficiency group. She found different patterns between the two groups. The advanced students were able to produce writing of similar complexity, accuracy, and fluency in both genres; however, the intermediate students showed less complex and accurate language in expository genres than in narratives. If a genre is too

difficult for a certain proficiency group, their use of language can be limited (e.g., less sophisticated vocabulary or less complex structures) because the difficulty they experience may overburden their working memory and increase the time they spend on revising or reviewing (Hayes, 2012; Kellogg, 1996). In Ruiz-Funes's study, possibly, high proficiency allowed the advanced learners to easily access the genre knowledge in their long-term memory, without overloading their working memory. Jeong (2017) also investigated genre effects and their interaction with proficiency (novice, intermediate, and advanced) in writing performance. She did not find significant differences between the two genres she tested, but she did find a significant interaction between genre and proficiency: Novice learners received higher scores on narratives than on expository essays, whereas advanced learners obtained higher scores on expository essays than on narratives. These studies' findings indicate the necessity of including proficiency in attempts to explain genre effects on the multifaceted aspects of writing.

2.3. Research questions

As discussed above, moving beyond a singular focus on assessing writing outcomes (e.g., the length of writing), this study investigates L2 learners' writing fluency-related behaviors and the cognitive processes behind them by exploring the effects of time constraints, genre, and proficiency. Drawing on Kellogg's (1996) model of writing, this study adopts a mixed-methods design and uses (a) keystroke logging to capture writing behaviors such as fluency, pausing, and revision, and (b) stimulated recall to reveal cognitive processes used by L2 learners (e.g., Révész, Kourtali, & Mazgutova, 2017; Van Waes & Leijten, 2013).

Most previous research regarding writing fluency behaviors and linguistic outcomes has

not yet touched upon the differences resulting from different time constraints. Rather, it has been conducted within short timed setting, possibly due to practicality and convenience, and inconsistent operations of time constraints in timed writing have been utilized. However, given that writing in extended time settings is widely preferable under certain circumstances such as classroom settings, extended timed writing should be investigated to increase ecological validity and reflect L2 writing in reality. Although this study cannot give the participants unlimited time for logistical reasons, it employs two timed conditions alternatively to remove the limitations of a time constraint as well as simulate untimed conditions. Based on Weigle's (2002) dimensions of time allowance, one condition gave 30 minutes, which has been widely used as a short time constraint for 300-word writing. The other condition doubled the time in which participants could complete the task.

Previous studies have indicated the impact of different genres and proficiency on L2 writing. In addition to the outcomes themselves, however, L2 learners' writing fluency behaviors underlying the writing process can provide further understanding of L2 writing. In particular, the extent to which genres and proficiency have impact on L2 writings may be different across linguistic outcomes and writing fluency behavior. Nevertheless, whether the observable traces of a person's cognitive activities such as pausing are due to differences in the demands of genres (e.g., Kellogg, 1990; Thorson, 2000), and the extent to which L2 learners show different writing processes depending on their L2 proficiency, have rarely been tested. To address this gap in research, this study delved into different L2 proficient learners' writing fluency behaviors and linguistic outcomes in different genres under different time constraints.

The study investigates the interwoven impact of time constraints, genre, and proficiency on L2 learners' writing fluency behavior to improve our understanding of L2 writing fluency.

Given the correlation between fluency and linguistic complexity and the impact of linguistic complexity on fluency, this study also explores linguistic complexity and writing quality, which ultimately provide relevant evidence for understanding L2 writing fluency. Moreover, by using keystroke logging software to explore these variables, the study employs a relatively innovative approach to assessing fluency.

The study addresses four specific research questions:

1. To what extent do proficiency and time constraints affect writing fluency behaviors and linguistic outcomes of L2 writers' writing in two genres?

2. As evidenced by the stimulated recall data, to what extent do proficiency and time constraints affect L2 writers' writing process in the two genres?

3. How do L2 proficiency and time constraints affect writing quality in two essay genres?

4. Which fluency measures are related to writing quality and linguistic complexity, and to what extent?

5. How do L2 writers perceive the effects of time constraints and genre on their writing?

CHAPTER 3. METHOD

3.1. Participants

The participants of the study were 128 EFL students (Age: $M = 22.75$, $SD = 2.31$; 38 males and 90 females) studying at a private university in Seoul, Republic of Korea, who all spoke Korean as a first language. The participants were selected according to three main criteria. First, they must have learned English as a second language. While they may have visited or resided in English-dominant countries, for instance in study-abroad programs, they must have learned English in instructional settings. 81 had never resided in an English-dominant country, while 47 had ($M = 6$ months, $SD = 15.13$ months). Second, they must have completed a required English for Academic Purposes class at their university. Third, they must have achieved high-intermediate or advanced proficiency according to standardized tests such as TOEFL or IELTS taken two years or less before the time of data collection. They received $25 for their participation, and the five who wrote the best essays, based on the essay scores, received additional compensation.

According to their standardized test scores, the participants were divided into intermediate (62 participants) and advanced (66 participants) groups.[1] However, because five participants' keystroke logging files had corruption errors, only 123 participants' data were included in the analysis, 60 in the intermediate group, and 63 in the advanced group (see Table 3).

---

[1] The high-intermediate participants had TOEFL scores of 72–94, TOEIC scores of 785–940, or IELTS scores of 5.5–6.5, each of which are equivalent, according to an ETS equivalency table, to Level B2 in the Common European Framework of Reference (CEFR) levels. The advanced participants had TOEFL scores of 95 or above, TOEIC Scores of 945 or above, or IELTS scores of 7 or above, which are equivalent to Level C1 (Papageorgiou, Tannenbaum, Bridgeman, & Cho, 2015).

Table 3.

Demographic Information of High Intermediate and Advanced Proficiency Students

		High intermediate (*N* = 60)	Advanced (*N*= 63)
Age, Mean (*SD*)		23.03 (*SD* = 2.05)	22.54 (*SD* = 2.54)
Gender	Male	16	20
	Female	44	43
Length of residence in English speaking countries, Mean (SD)		3.07 months (*SD* = 6.73)	8.86 months (*SD* = 20)

3.2. Materials

A narrative writing prompt and an argumentative writing prompt were used to investigate genre effects on the participants' writing. In order to minimize potential topic effects, the topics were controlled by using the prompts on the same theme, learning a foreign language. They came from Yoon (2017; see Appendix A).

In order to ensure intergroup comparability in terms of proficiency levels, a cloze test was administered to measure the L2 learners' English proficiency at the time of data collection (Appendix B). The cloze test was used because it is considered to be a valid measure of global proficiency when the focus of research is related to literacy skills (Wu & Ortega, 2013). The test was composed of 50 items, and the L2 learners were asked to finish it within 25 minutes. The cloze test was scored by the acceptable answer scoring method, which considers all contextually acceptable answers as correct answers and, consequently, increases test reliability. Correct

answers received one point; thus, the scores could range from 0 to 50. The results of the cloze test were found to be reliable (Cronbach's $\alpha = .79$), suggesting its consistency in distinguishing the participants.

A timed key-boarding skill test (Appendix C) was used to ensure the comparability of the groups in terms of typing speed, which might affect their fluency in writing (Barkaoui, 2016). The participants were asked to copy a sentence as many times as they could in two minutes. By calculating their typing speed measured by the number of total characters typed, the study was able to control for typing speed when assessing writing fluency. The typing speed in each group was compared to ensure intergroup comparability.

Two questionnaires were used. First, the Language Experience and Proficiency Questionnaire (Appendix D) developed by Marian, Blumenfeld, and Kaushanskaya (2007) was used to collect the participants' biographic information including age, sex, length of residence in English-dominant countries, and standardized English test scores. In addition, an exit questionnaire (Appendix E) adapted from one employed by Yoon (2017) was used to ask the participants' perceptions of time constraints and genres. The questionnaire was composed of two open-ended questions and eight items to be rated on a nine-point Likert scale.

In order to measure the quality of the participants' argumentative essays, the analytic rubric provided in Connor-Linton and Polio (2014) was used (Appendix G). Because this analytic rubric can provide detailed information on various aspects of L2 writers' performance, it is preferable to a holistic rubric (Weigle, 2002). The rubric is an adapted version of the ESL composition profile (Jacobs, Zinkgrap, Wormuth, Hartfiel, & Hughey, 1981) that is most widely used, and the full score is 90 points. It consists of five subscales (content, organization, vocabulary, language use, and mechanics); the full score of each of the first four subscales is 20

points, and the full score of the mechanics subscale is 10 points.

The rubric was designed for assessing argumentative essays, but the current study also required an analytic rubric for narrative essays. I therefore revised the rubric to make it applicable to narratives (Appendix H). Following Polio and Lim (under review), two expert raters were given three narratives on the same topic and told to rank the essays in terms of quality while talking about their rankings. Both raters were doctoral students in second language studies who had taught ESL and EFL students and rated essays when working at an English language center. One rater was an experienced English teacher and the other an IELTS certified examiner. I asked the raters to rank the essays only by quality, and I audiorecorded their descriptions. Both gave the same ranks of ratings to the three essays, and described the quality of the narratives. After they rated and discussed the quality of the narratives, I gave them the analytic rubric for argumentative essays from Connor-Linton and Polio (2014) and asked them to rate the narratives based on the rubric. They discussed some difficulties of rating narratives with the argumentative essay rubric and suggested possible ways to adapt it for narratives. Based on their discussion, I revised the rubric. The validity of the revised rubric was then confirmed by an L2 writing expert, a professor who has conducted research on L2 writing over 25 years at a university in the United States.

3.3. Procedures

The experimental design was mixed, with one within-subject and two between-subject factors. The independent variables were genre (within-subject), timing conditions (between-subject), and proficiency (between-subject). The dependent variables were syntactic complexity,

fluency behaviors, and writing quality. I met with each participant individually in a conference room at their school on two separate days. Each day, the participants were asked to write one 300-word essay on a computer. The participants were not allowed to use reference materials or other resources to complete the essays. Their writing was recorded by Inputlog 7.0 (Leijten & Van Waes, 2013), a keystroke logging program. Half of the participants were assigned randomly to the shorter time group and half to the longer time group. They were given 30 minutes in the short-timed condition and 60 minutes—double the time to mimic an untimed condition—in the long-timed condition. Giving the students unlimited time was impossible for logistical reasons; doubling the time was an attempt to remove the limitations of a time constraint.

Each participant was randomly assigned either the narrative or the argumentative essay prompt on the first day and the other on the second day, in order to counterbalance the order of the genres. To minimize testing effects from a repeated design, the participants were asked to schedule the second day of the experiment at least a week after the first day. On the first day, they completed the cloze test and the background questionnaire right after they finished their writing (either narrative or argumentative). On the second day, they completed the timed key-boarding skill test and the exit questionnaire after finishing their writing (either narrative or argumentative).

A total of 16 participants (eight each day, with one from each proficiency group, in each time condition, and after writing in each genre type; see Table 4) were randomly selected for stimulated-recall sessions in order to triangulate the data. Stimulated recall is useful for understanding the participants' thoughts on their writing process. Previous research has used stimulated-recall protocols to better understand the process of writing in terms of what participants pay attention to, the difficulties they encounter, and the online behaviors they show

(Barkaoui, 2015; Lindgren, 2005). This study's stimulated-recall protocols followed those

suggested by Gass and Mackey (2017) and Barkaoui (2015). The stimulated-recall session took

approximately an hour, and the selected participants completed the session right after they

finished their writing of the day, before completing the other tests. The participant and the

researcher watched the screen recording generated by Camtasia together; the participant was told

to pause at any time to comment. The researcher also stopped the recording whenever the

participant paused or revised. If the participants could not recall their writing behaviors, further

questions were not asked (Appendix F). To elicit rich data, the stimulated-recall sessions were

conducted in their L1, Korean.

Table 4

Participants

English proficiency	Timing constraints	Genres (two different days)	Stimulated recalls
High-intermediate ($N = 60$)	Short-timed ($N = 30$) Long-timed ($N = 30$)	Narrative Argumentative	Two participants at two different proficiency levels conducted the sessions after they finished 30 min/60 min narratives and
Advanced ($N = 63$)	Short-timed ($N = 33$) Long-timed ($N = 30$)	Narrative Argumentative	argumentative essays

Table 5

Cloze Test Scores

Conditions	Short-timed (30 minutes)		Long-timed (60 minutes)	
	M (*SD*)	95% CI	*M* (*SD*)	95% CI
High-intermediate	30.03 (4.73)	28.27, 31.80	28.07 (5.50)	26.01, 30.12
Advanced	37.00 (4.76)	35.31, 38.69	36.77 (5.36)	34.76, 38.77

Note. Total score is 50.

Table 6

Keyboarding Skill Test Scores (Number of Total Characters Typed within 2 Minutes)

Conditions	Short-timed (30 minutes)		Long-timed (60 minutes)	
	M (*SD*)	95% CI	*M* (*SD*)	95% CI
High-intermediate	509.17 (111.02)	467.71, 550.62	493.43 (130.90)	444.56, 542.31
Advanced	621.48 (78.92)	593.50, 649.47	579.57 (93.14)	544.79, 614.35

3.4. Scoring

Table 5 presents the descriptive statistics of the groups' scores. To ensure group comparability, independent samples *t*-tests were performed. A statistical difference between high-intermediate and advanced proficiency levels was found ($t(121) = -8.52$, $p < .001$, 95% CI $= [-9.66, -6.01]$). For the short-timed and long-timed group comparisons, no statistical differences were found within the high-intermediate proficiency group ($t(58) = 1.49$, $p = .14$, 95% CI $= [-.69, 4.62]$) or the advanced proficiency group ($t(61) = .18$, $p = .90$, 95% CI $= [-2.32$,

2.78]).

For the key boarding test, independent samples t-tests were performed to find the comparability of the groups (see Table 6). The results showed a significant difference between proficiency levels ($t(107) = -5.24$, $p < .001$, 95% CI = [−138.08, −62.36]), but no significant difference between time constraint conditions ($t(121) = 1.51$, $p = .13$, 95% CI = [−9.76, 72.55]). For the short-timed and long-timed group comparisons, no significant differences were found within the high-intermediate group ($t(58) = .502$, $p = .62$, 95% CI = [−46.99, 78.46]) or the advanced group ($t(61) = 1.93$, $p = .06$, 95% CI = [−1.45, 85.29]). Therefore, the keyboarding skills that may affect writing fluency behaviors differed between the two proficiency levels but were similar in the two time constraint groups.

Two native English speakers, who were expert raters and had taught ESL and EFL students, rated the essays based on the rubrics. Both raters were instructors at an English language center at a university and were studying towards their master's degrees in TESOL. The raters were trained in a two-hour norming session where they rated sample narratives and argumentative essays that were not part of this study and discussed their scoring. If a discrepancy in any subscale was greater than two points, the raters resolved the discrepancy through discussion. After the norming session, the raters independently rated all of the essays, and the average scores obtained from the two raters were used for the analysis. If some essays received discrepant scores (subscale scores differing by three or more), a third rater rated the essays, and the two closer scores were utilized to find average scores. Because the prompts and rubrics were different for the two genres, interrater reliability was calculated by genre. The interrater reliability of the total scores for the narratives was $r = .81$ (content: $r = .74$, organization: $r = .71$, vocabulary: $r = .70$, language use: $r = .74$, and mechanics: $r = .77$). The interrater reliability of

37

the total scores for the argumentative essays was $r = .85$ (content: $r = .75$, organization: $r = .76$, vocabulary: $r = .75$, language use: $r = .78$, and mechanics: $r = .82$). According to Brown, Glasswell, and Harland (2004), reliability of 0.70 is a benchmark for structured rubrics, and thus the interrater reliability for both the narrative and argumentative essays is within an acceptable range.

3.5. Analysis

To analyze the syntactic complexity of the participants' written texts, the 14 syntactic complexity measures in Lu's (2010) syntactic complexity analyzer were used. Based on previous studies (Lu, 2011; Yoon & Polio, 2017), some inaccurate measures for development and genre effects such as clauses per sentence (C/S), complex T-unit ratio (CT/T), and sentence coordination ratio (T/S) were excluded. For lexical complexity, the *D* index and the lexical sophistication measure (the logarithm of word frequency for all words and average length of word) were calculated by using Coh-Metrix (McNamara, Graesser, McCarthy, & Cai, 2014). From among the measures of the frequency of all words, the logarithm of word frequency for all words (WF) and average word length (WL) were selected in order to prevent rare words from creating a limiting factor. In interpreting WF, a lower value means less frequent words and a higher value means more frequent words. Following Yu (2010), spelling mistakes were corrected before running the syntactic complexity analyzer and Coh-Metrix.

Table 7

Fluency Measures (Adapted from Van Waes & Leijten, 2015)

Measures	Definitions
Process	Number of words produced, including deleted words
Product	Number of words produced in the final text
Ratio of process and product	Proportion between process and product measures
P-burst	A string of actions delimited by an initial pause and end pause exceeding the defined pause threshold (2000 ms).
R-burst	Language bursts that were bounded by a revision.

#Id	Event Type	Output	Position	DocLength	Character Production	StartTime	StartClock	EndTime	EndClock	ActionTime	PauseTime	PauseLocation	X	Y
0	focus	Wordlog_2004_20180416200857 - Microsoft Word			0	37705	00:00:37	37705	00:00:37	0	0	CHANGE		
2	keyboard	L	0	1	1	38003	00:00:38	38094	00:00:38	389	0	BEFORE SENTENCES		
3	keyboard	e	1	2	2	40067	00:00:40	40265	00:00:40	198	2064	WITHIN WORDS		
4	keyboard	a	2	3	3	40261	00:00:40	40403	00:00:40	142	194	WITHIN WORDS		
5	keyboard	r	3	4	4	40523	00:00:40	40604	00:00:40	81	262	WITHIN WORDS		
6	keyboard	n	4	5	5	40687	00:00:40	40853	00:00:40	166	164	WITHIN WORDS		
7	keyboard	i	5	6	6	40849	00:00:40	41043	00:00:41	194	162	WITHIN WORDS		
8	keyboard	n	6	7	7	41021	00:00:41	41116	00:00:41	95	172	WITHIN WORDS		
9	keyboard	g	7	8	8	41258	00:00:41	41330	00:00:41	72	237	WITHIN WORDS		
10	keyboard	SPACE	8	9	9	42259	00:00:42	42373	00:00:42	114	1001	AFTER WORDS		
11	keyboard	a	9	10	10	42423	00:00:42	42565	00:00:42	142	164	BEFORE WORDS		
12	keyboard	SPACE	10	11	11	42568	00:00:42	42647	00:00:42	79	145	AFTER WORDS		
13	keyboard	l	11	12	12	42769	00:00:42	42873	00:00:42	104	201	BEFORE WORDS		
14	keyboard	a	12	13	13	42953	00:00:42	43054	00:00:43	101	184	WITHIN WORDS		
15	keyboard	n	13	14	14	43359	00:00:43	43421	00:00:43	62	406	WITHIN WORDS		
16	keyboard	g	14	15	15	43977	00:00:43	44079	00:00:44	102	618	WITHIN WORDS		
17	keyboard	u	15	16	16	44202	00:00:44	44316	00:00:44	114	225	WITHIN WORDS		
18	keyboard	a	16	17	17	44388	00:00:44	44480	00:00:44	92	186	WITHIN WORDS		
19	keyboard	BACK	17	18	18	45855	00:00:45	45926	00:00:45	71	1467	REVISION		
20	keyboard	BACK	16	17	18	46249	00:00:46	46602	00:00:46	353	394	REVISION		
21	keyboard	BACK	15	16	18	46724	00:00:46	46796	00:00:46	72	475	REVISION		
22	keyboard	BACK	14	15	18	46927	00:00:46	46978	00:00:46	51	203	REVISION		
23	keyboard	BACK	13	14	18	47109	00:00:47	47150	00:00:47	41	182	REVISION		

Figure 2. Inputlog 7.0: Screen capture

To analyze fluency, the data recorded by Inputlog 7.0 was used (see Figure 2). Following

Van Waes and Leijten (2015), several measures were calculated: in the writing product, words

per minute; in the writing process, words per minute, number of P-bursts, mean typed characters

in P-bursts (P-burst length), number of pauses within words, number of pauses between words/sentences/paragraphs, number of R-bursts, mean typed characters in R-bursts (R-burst length), and the ratio of process and product (proportion between product and process measures). The number of characters per minute in the writing process includes the number of characters that the learners deleted in the writing process whereas the number of characters per minute in the writing product only considers the number of characters in the final product. The number of pauses within words is related to the efficiency of typing, word finding and spelling behaviors (Torrance & Galbraith, 2006). The number of pauses between words is usually caused by lexical retrieval and editing process whereas the number of pauses between clauses can include planning processes (Wengelin, 2006). The number of pauses between sentences or paragraphs is likely to be associated with planning processes (Wengelin, 2006). This fluency analysis identified bursts, which are sequences of keystrokes without long pauses. Thus, a burst is a chunk of words that is bounded by breaks in written production. Bursts are therefore a useful measurement to show efficiency in writing (Chenoweth & Hayes, 2003). According to Chenoweth and Hayes (2003), P-bursts are defined as the bursts bounded by pausing followed by continued written production. R-bursts are defined as the bursts bounded by revision of the language produced during the burst. More fluent writers can take fewer pauses than less fluent writers. Consequently, more fluent writers may show a lower number of P-bursts than less fluent writers. In addition, more fluent writers can write more words between pauses and show longer lengths of P-bursts than less fluent writers. Following the previous studies, the threshold for pauses was set to 2000 milliseconds (Spelman Miller et al., 2008; Van Waes & Leijten, 2015); in other words, only pauses over 2000 milliseconds were counted. Table 7 provides explanations on the fluency measures.

Table 8

Coding Categories (Adapted from Révész et al., 2017)

Process/Subprocess		Example comments (English translation)
Planning	Content	I was thinking about two things. The first one was the process of learning English when I lived in the States. The other one was the process of learning Chinese in high school, when I did not acquire much because I was too old to learn quickly. I was thinking what to say here.
	Organization	I was thinking about the whole structure of this writing. How can I connect this paragraph to the next one? How can I connect this sentence and paragraph to the whole writing? I was thinking these things.
Translation	Lexical retrieval	Because the next sentence is a fact that is hard to generalize, I was thinking about using different and more sophisticated words instead of saying "act positively."
	Syntactic encoding	I was writing this part, "translation services currently provided by." I stopped and asked myself if the verb *provide* should take an object. I just wrote "currently being provided by" and ended the sentence. But I felt that I was wrong. The verb, *provide* needs a provider. I was thinking about whether *provide* needs an object.
	Cohesion	As I was looking at this word, *therefore* does not fit in here. It might be better to use *because* in order to change this sentence.
	Unspecified	I found this part awkward. I wanted to make this sentence more natural.
Monitoring		I was looking at this sentence. From now on, I was skimming from the beginning and correcting some mistakes.

3.5.1. Qualitative analysis

With respect to the stimulated-recall data, following Kellogg's (1996) model and Révész et al. (2017), the participants' comments were transcribed verbatim and coded into three categories - planning, translation, and monitoring – as shown in Table 8 - using MAXQDA. With

respect to pausing and revision comments, following Stevenson et al. (2006), their comments were counted by pause location and type of revision, as determined by watching the video generated by Camtasia. Their comments about pausing and revision were calculated in terms of pause location and types of revision. Three participants' comments (about 18 percent of the data) were double-coded to check intercoder agreement reliability (95%), and any discrepancy was resolved through discussion.

3.5.2. Statistical analysis

In order to address the research questions, the complexity indices, fluency indices, and writing quality of the participants' essays were analyzed. SPSS 25 was used to determine whether there were statistical differences between the genres, the timing conditions, and the proficiency levels in terms of complexity, fluency, and writing quality. With the explore function in SPSS, descriptive statistics and 95% confidence intervals were obtained. Multicollinearity was controlled between measures ($r > .90$). If the two measures were multicollinear, only one of the measures was included in the analysis. Because of multicollinearity, some syntactic complexity measures (mean length of T-unit, clauses per T-unit, dependent clauses per T-unit, coordinate clauses per T-unit, complex nominals per clause) were excluded. Table 9 summarizes the measures included in the analysis.

Table 9

Linguistic Measures as Dependent Variables

	Complexity and fluency measures
Length of production	Mean length of sentence (MLS)
	Mean length of clause (MLC)
Subordination	Dependent clause ratio (DC/C)
Coordination	Coordinate phrases per clause (CP/C)
Particular structure	Complex nominals per T-unit (CN/T)
	Verb phrases per T-unit (VP/T)
Lexical sophistication	The logarithm of word frequency for all words (WF)
	Mean length of word (WL)
Lexical diversity	D
Fluency	Process: Words per minute
	Product: Words per minute
Pausing	The number of P-bursts
	The mean typed characters per P-burst (P-burst length)
	The number of pauses within words
	The number of pauses between words
	The number of pauses between sentences
	The number of pauses between paragraphs
Revision	The ratio of process and product
	The number of R-bursts
	The mean typed characters per R-burst (R-burst length)

The independent variables were genre (within-subject), timing conditions (between-subject), and proficiency (between-subject). The dependent variables were linguistic complexity, fluency behaviors, and writing quality. Regarding the first research question, in order to examine the effect of genres, timing conditions, and proficiency levels on the dependent variables (complexity and fluency), a repeated-measures multivariate analysis of variance (MANOVA) was conducted. Every student wrote essays in two genres, and did so in one of the two time constraint conditions, and the student's proficiency was either high-intermediate or advanced. Because the prompts and rubric are different in the two genres, text quality (writing scores) for the genres was not included in the MANOVA. Evaluation of the homogeneity of variance-covariance matrices (Box's M), error variances (Levene's test), linearity, non-multicollinearity, and normality assumptions underlying MANOVA did not reveal any substantial anomalies. Given the number of comparisons, the a priori alpha level was set at $p < .0025$ with Bonferroni adjustment (.05/20).

For the second research question about the effect of time constraints and proficiency on text quality, a two-way analysis of variance (ANOVA) for each genre was conducted. Given the multiple comparisons, the a priori alpha level was set at $p < .0083$ with Bonferroni adjustment (.05/6). With regard to the third research question, to explore the relationship between writing fluency measures and writing quality and to determine which fluency measures predict writing quality, a correlation and multiple regression analysis were performed. For the fourth research question about the L2 learners' perceptions, to analyze the results of the questionnaire on genre and time constraints, one-way analyses of variance (ANOVA) and a post-hoc Bonferroni test were used to look for differences in the learners' perceptions of their writing tasks. The a priori alpha level was set at $p < .0062$ with Bonferroni adjustment (.05/8).

Along with exact *p*-values, effect sizes for inferential statistics (Cohen's *d*) are reported. Cohen's *d* is considered to be the most appropriate effect size estimate. The effect size can tell the magnitude of quantitative findings and observed differences between two conditions in standard deviation units (Norris & Ortega, 2000; Plonsky & Oswald, 2014). According to Plonsky and Oswald (2014), small, medium, and large effect sizes of Cohen's *d* correspond to values of .40, .70, and 1, respectively.

CHAPTER 4. RESULTS

4.1. Quantitative analysis

The descriptive statistics for writing fluency behaviors and linguistic outcomes by time constraints, proficiency, and genres are presented in Table 10. The learners in each group wrote the narrative essays and the argumentative essays on two different days. Although the 95% confidence intervals for the four groups overlap, there seem to be differences between the groups. Within groups, the two genres differed in terms of syntactic complexity, fluency, and writing fluency behaviors (pausing and revision). The L2 learners tended to produce more complex language, such as higher syntactic complexity and lexical complexity and less fluent writing behaviors, such as shorter P-burst lengths in argumentative essays than narratives. The short-timed groups showed longer P-burst lengths than the long-timed groups. The advanced students tended to show higher syntactic complexity (i.e., MLS, MLC, CN/T and VP/T) and fluency (i.e., process: words per minutes and product: words per minute) than the high-intermediate students.

Table 10

Descriptive Statistics: Writing Fluency Behaviors and Linguistic Outcomes by Time Constraints, Proficiency, and Genres

Measures	High-intermediate short-timed (N = 30)				High-intermediate long-timed (N = 30)				Advanced short-timed (N = 33)				Advanced long-timed (N = 30)			
	Nar		Arg		Nar		Arg		Nar		Arg		Nar		Arg	
	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI
MLS	17.96 (5.31)	15.98, 19.94	18.71 (6.01)	16.47, 20.96	16.43 (3.77)	15.02, 17.84	17.60 (3.90)	16.14, 19.05	20.83 (4.57)	19.20, 22.45	21.04 (4.59)	19.41, 22.66	19.77 (4.14)	18.22, 21.32	20.75 (3.40)	19.48, 22.02
MLC	8.83 (1.45)	8.29, 9.37	9.37 (1.31)	8.89, 9.86	8.35 (1.02)	7.97, 8.74	9.35 (1.36)	8.84, 9.86	9.33 (1.27)	8.88, 9.78	10.60 (1.62)	10.03, 11.18	9.07 (1.53)	8.50, 9.64	10.27 (1.24)	9.81, 10.73
DC/C	.39 (.10)	.35, .43	.39 (.12)	.35, .43	.40 (.08)	.37, .42	.37 (.05)	.35, .39	.44 (.10)	.40, .47	.41 (.10)	.37, .44	.42 (.08)	.39, .45	.43 (.09)	.40, .46
CP/C	.18 (.09)	.14, .20	.18 (.10)	.14, .21	.20 (.07)	.18, .24	.23 (.11)	.19, .27	.24 (.12)	.20, .28	.27 (.15)	.22, .32	.21 (.10)	.18, .25	.26 (.10)	.22, .30
CN/T	1.69 (.57)	1.49, 1.89	2.25 (.65)	2.01, 2.49	1.56 (.46)	1.39, 1.73	1.93 (.48)	1.75, 2.11	1.98 (.69)	1.74, 2.22	2.50 (.65)	2.27, 2.73	1.89 (.52)	1.69, 2.08	2.52 (.63)	2.28, 2.75
VP/T	2.46 (.45)	2.29, 2.62	2.50 (.69)	2.23, 2.75	2.31 (.41)	2.16, 2.49	2.35 (.37)	2.21, 2.49	2.76 (.68)	2.52, 3.00	2.70 (.68)	2.46, 2.94	2.65 (.43)	2.49, 2.81	2.70 (.44)	2.54, 2.86
WL	1.47 (.08)	1.44, 1.50	1.60 (.11)	1.56, 1.64	1.46 (.06)	1.44, 1.48	1.59 (.07)	1.57, 1.62	1.49 (.06)	1.47, 1.51	1.65 (.08)	1.62, 1.68	1.47 (.09)	1.44, 1.50	1.62 (.07)	1.60, 1.66
WF	3.09 (.09)	3.05, 3.12	3.03 (.10)	2.99, 3.06	3.09 (.07)	3.07, 3.12	3.05 (.07)	3.02, 3.07	3.07 (.07)	3.04, 3.09	3.01 (.09)	2.98, 3.04	3.10 (0.07)	3.07, 3.12	2.96 (.23)	2.88, 3.05
D	87.77 (17.29)	81.31, 94.22	83.86 (18.93)	76.78, 90.93	87.66 (15.07)	82.04, 93.29	84.98 (16.40)	78.85, 91.10	93.93 (17.45)	87.75, 100.12	88.32 (19.09)	81.55, 95.09	88.98 (13.05)	84.11, 93.86	88.22 (13.65)	83.13, 93.32

Table 10 (cont'd)

Measures	High-intermediate short-timed (N = 30)				High-intermediate long-timed (N = 30)				Advanced short-timed (N = 33)				Advanced long-timed (N = 30)			
	Nar		Arg		Nar		Arg		Nar		Arg		Nar		Arg	
	M (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI
Process: Words per minute	17.62 (4.97)	15.77, 19.48	15.58 (4.20)	14.01, 17.15	11.83 (2.86)	10.77, 12.90	10.83 (3.05)	9.70, 11.97	20.73 (4.69)	19.07, 22.40	18.12 (4.19)	16.63, 19.60	17.15 (4.75)	15.38, 18.92	15.64 (4.60)	13.92, 17.36
Product: Words per minute	12.88 (3.94)	11.41, 14.35	11.16 (3.16)	9.97, 12.33	8.00 (2.31)	7.13, 8.86	7.05 (1.97)	6.30, 7.78	14.57 (4.55)	12.96, 16.18	12.40 (3.10)	11.30, 13.49	11.56 (3.86)	10.12, 13.00	10.38 (3.76)	8.98, 11.79
Number of P-bursts	3.26 (1.06)	2.86, 3.66	3.48 (.93)	3.13, 3.82	3.72 (.86)	3.40, 4.04	3.78 (.72)	3.51, 4.05	3.27 (.81)	2.98, 3.55	3.50 (.79)	3.22, 3.78	3.39 (.71)	3.13, 3.65	3.44 (.61)	3.21, 3.67
P-burst length	41.39 (27.48)	31.13, 51.65	33.05 (16.04)	27.06, 39.04	22.07 (9.65)	18.4, 25.67	20.22 (7.95)	17.25, 23.20	43.62 (17.82)	37.30, 49.94	37.51 (15.05)	32.18, 42.85	34.33 (17.34)	27.85, 40.80	31.92 (12.50)	27.25, 36.59
Pause within words	.29 (.15)	.23, .34	.46 (.42)	.30, .62	.34 (.21)	.26, .42	.27 (.54)	.07, .48	.37 (.35)	.25, .50	.37 (.28)	.27, .47	.31 (.24)	.22, .39	.32 (.32)	.20, .44
Pause between words	1.69 (.72)	1.42, 1.96	1.73 (.83)	1.42, 2.04	1.68 (.60)	1.45, 1.90	1.13 (.71)	.87, 1.40	1.52 (.59)	1.31, 1.73	2.24 (.95)	1.90, 2.58	1.57 (.54)	1.37, 1.77	1.20 (.77)	.91, 1.48

Table 10 (cont'd)

Measures	High-intermediate short-timed (N = 30)				High-intermediate long-timed (N = 30)				Advanced short-timed (N = 33)				Advanced long-timed (N = 30)			
	Nar		Arg		Nar		Arg		Nar		Arg		Nar		Arg	
	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI
Pause between sentences	.19 (.15)	.14, .25	.17 (.13)	.12, .22	.15 (.11)	.11, .19	.10 (.08)	.07, .13	.17 (.13)	.12, .22	.19 (.10)	.15, .22	.19 (.17)	.13, .26	.15 (.13)	.10, .20
Pause between paragraphs	.05 (.05)	.02, .06	.05 (.06)	.02, .07	.03 (.03)	.02, .05	.03 (.03)	.02, .04	.05 (.05)	.02, .06	.06 (.06)	.04, .08	.03 (.03)	.02, .04	.04 (.04)	.02, .05
Ratio of process and product	.68 (.10)	.64, .71	.69 (.09)	.65, .72	.63 (.13)	.59, .68	.61 (.13)	.56, .66	.66 (.13)	.62, .71	.65 (.10)	.62, .69	.65 (.09)	.61, .68	.63 (.10)	.59, .67
Number of R-bursts	6.32 (7.15)	3.65, 8.99	5.50 (4.72)	3.74, 7.26	3.67 (2.73)	2.65, 4.69	3.30 (2.78)	2.26, 4.34	7.38 (5.43)	5.46, 9.31	5.60 (5.12)	3.79, 7.42	4.67 (3.93)	3.20, 6.14	2.85 (3.06)	1.71, 3.99
R-burst length	11.71 (5.82)	9.53, 13.88	11.59 (5.88)	9.39, 13.78	10.83 (5.99)	8.59, 13.07	10.27 (4.11)	8.73, 11.80	12.49 (5.02)	10.70, 14.27	10.90 (4.26)	9.40, 12.42	12.69 (4.45)	11.03, 14.35	12.78 (4.66)	11.04, 14.52

A repeated measures MANOVA was performed, using 20 dependent measures to analyze within genres (within-subject variable). The independent variables were proficiency and time constraints. The MANOVA indicated statistically significant genre differences of the combined dependent variables according to Wilks' Lambda (.169; $F(20, 100) = 24.508$, $p = < .001$, $d = .90$). As shown in Table 11, follow-up univariate ANOVAs found statistically significant differences between the two genres in MLC ($p = < .001$), CN/T ($p = < .001$), WL ($p = < .001$), WF ($p = < .001$), D ($p = < .001$), process: words per minute ($p = < .001$), product: words per minutes ($p = < .001$), P-burst length ($p = < .001$), and the number of R-bursts ($p = .001$). A comparison of effect sizes suggested that genre differences had the greatest, though still moderate, effect on complexity, fluency, and writing fluency behaviors. In addition, the interaction between genre and time (.738, $F(20, 100) = 1.771$, $p = .034$, $d = .24$) was found to be statistically significant, indicating that the effect of genre on the linguistic measures was not the same in the two time constraint conditions. This result suggests that the learners wrote differently in the two genres depending on the given time. In contrast, according to Wilks' Lambda, there was no interaction between genre and proficiency (.804, $F(20, 100) = 1.220$, $p = .254$, $d = .19$), suggesting that the high-intermediate learners and advanced learners constructed their writing in similar ways regardless of genre.

Univariate testing showed the interaction between genre and time to be significant in the number of pauses between words ($F(1, 119) = 18.764$, $p = < .001$, $d = .78$). Figure 3 shows that the participants in the short-timed group made fewer pauses between words in narratives than in argumentative essays; however, the participants in the long-timed group made fewer pauses between words in argumentative essays than in narratives.
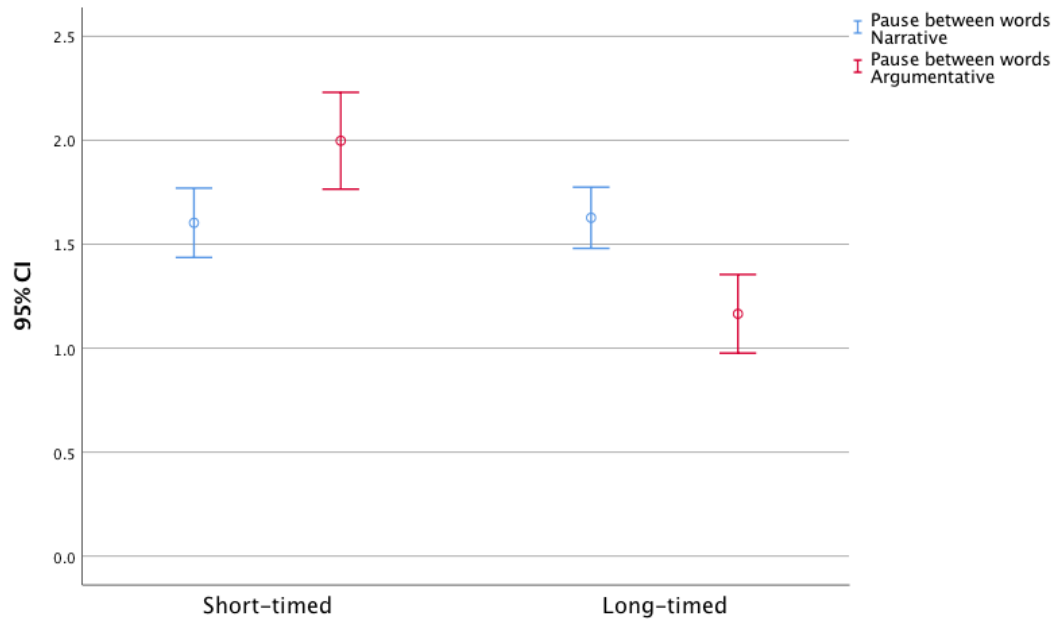
Figure 3. Means of pauses between words in the two genres.

A close examination of the results of the follow-up univariate ANOVAs and the descriptive statistics shows that the patterns of differences were different for each measure. The argumentative genre led the participants to produce higher MLC, CN/T, WL, and WF than did the narrative genre. As Figures 4, 5, 6, and 7 demonstrate, the argumentative genre elicited more complex language than the narrative genre across the groups. The narrative genre showed higher P-burst lengths, product: words per minutes and number of R-bursts than the argumentative genre. As Figures 8, 9, and 10 show, when the participants wrote narratives, they showed more fluent writing behaviors than when they wrote argumentative genres. In sum, although the MANOVA detected differences in the genres, the follow-up analyses showed that the patterns of genre differences varied for each measure.

Table 11

Repeated Measures MANOVA: Effects of Time Constraints and Proficiency on Writing Fluency

Behaviors and Linguistic Outcomes within Genres

Measures	Genre			Genre * Proficiency			Genre * Time		
	F	P	d	F	p	d	F	p	d
MLS	7.219	.008	.48	.402	.527	.11	1.046	.309	.18
MLC	70.817	<.001*	1.51	3.820	.053	.35	.621	.432	.14
DC/C	2.364	.127	.27	.078	.781	.05	.008	.931	.02
CP/C	5.456	.021	.42	1.608	.207	.23	.795	.374	.16
CN/T	119.154	<.001*	1.96	1.345	.248	.20	.155	.695	.07
VP/T	.128	.722	.06	.222	.638	.09	.392	.533	.13
WL	352.196	<.001*	3.38	2.678	.104	.30	.000	.985	0
WF	35.823	<.001*	1.07	2.848	.094	.30	1.926	.168	.25
D	4.227	.042	.37	.001	.972	0	.924	.338	.17
Process: Words per minute	43.234	<.001*	1.18	.994	.321	.18	3.890	.051	.35
Product: Words per minute	33.893	<.001*	1.05	.416	.520	.11	2.951	.088	.31
Number of P-bursts	6.398	.013	.45	.006	.940	.01	2.564	.112	.29
P-burst length	14.094	<.001*	.67	.113	.737	.06	4.193	.043	.37
Pause within words	.422	.517	.11	.271	.604	.09	1.562	.214	.23
Pause between words	.185	.668	.08	4.773	.031	.39	18.764	<.001*	.78
Pause between sentences	2.092	.151	.26	.660	.418	.15	1.877	.173	.24
Pause between paragraphs	.086	.770	.05	1.198	.276	.06	.602	.439	.14
Ratio of process and product	1.829	.179	.05	.123	.726	.06	1.932	.167	.25
Number of R-bursts	12.662	.001*	.64	3.200	.076	.32	.093	.761	.05
R-burst length	1.61	.206	.22	.228	.634	.09	.517	.474	.13
Wilk's Lambda	25.182	<.001*	.90	1.220	.254	.19	1.968	.015*	.25

* $p < .0025$ (Bonferroni adjustment for dependent variables)
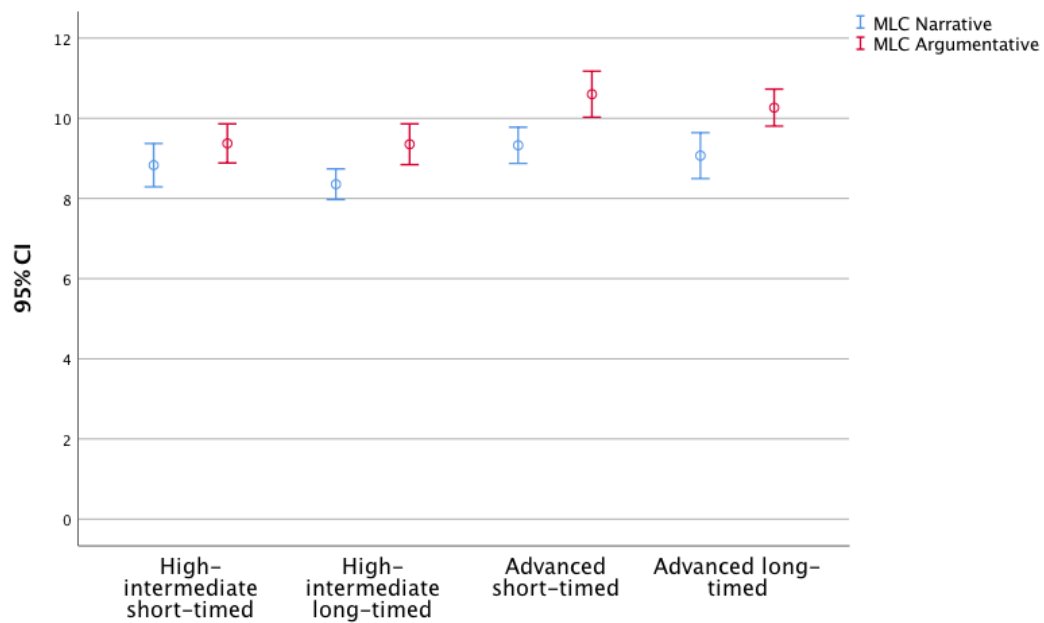
Figure 4. Genre differences in MLC
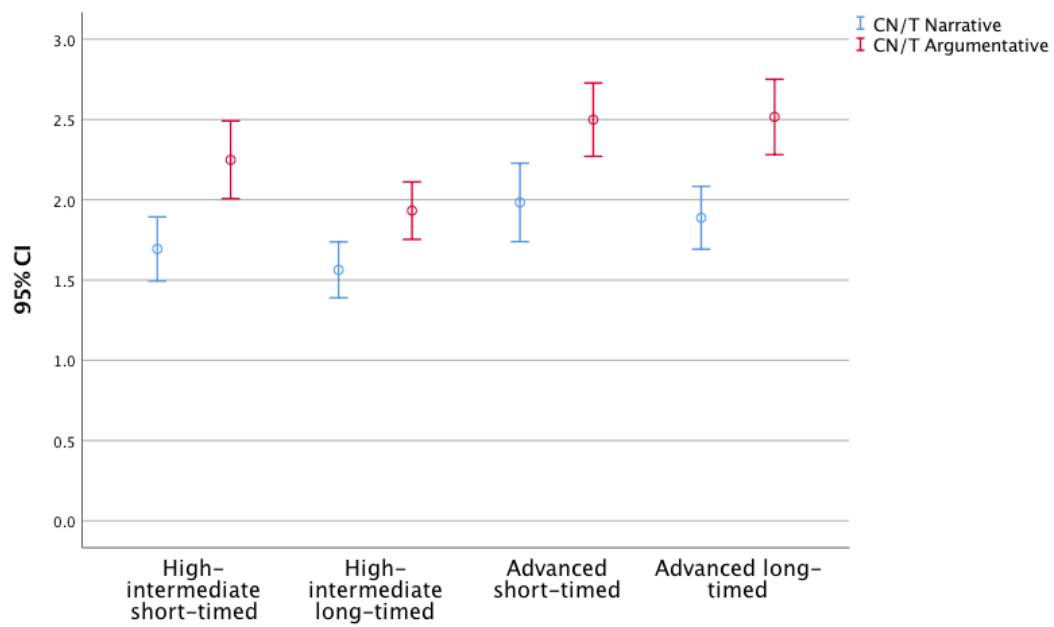
Figure 5. Genre differences in CN/T
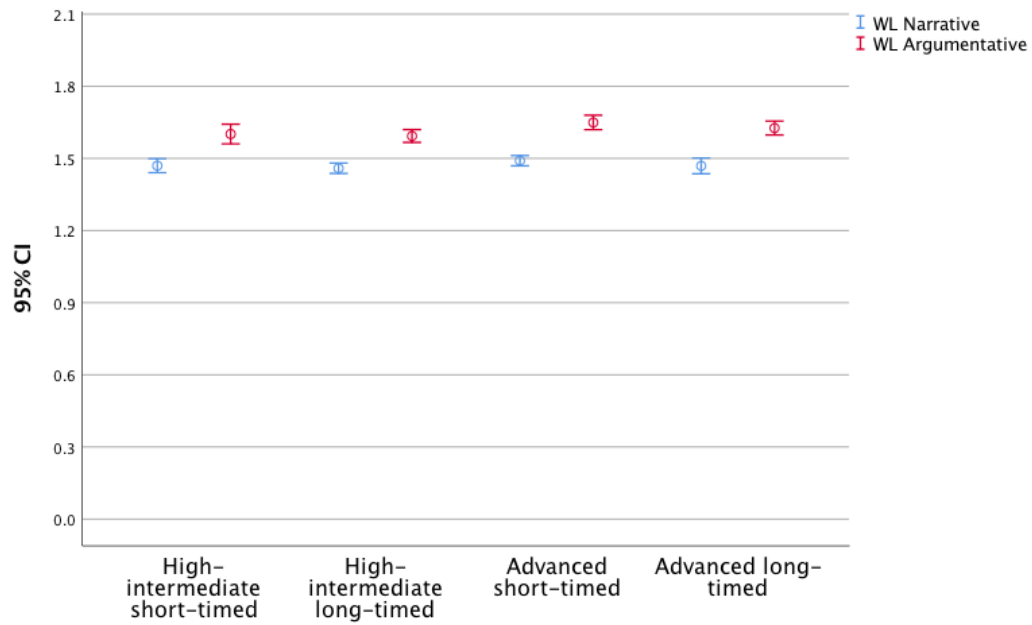
Figure 6. Genre differences in WL
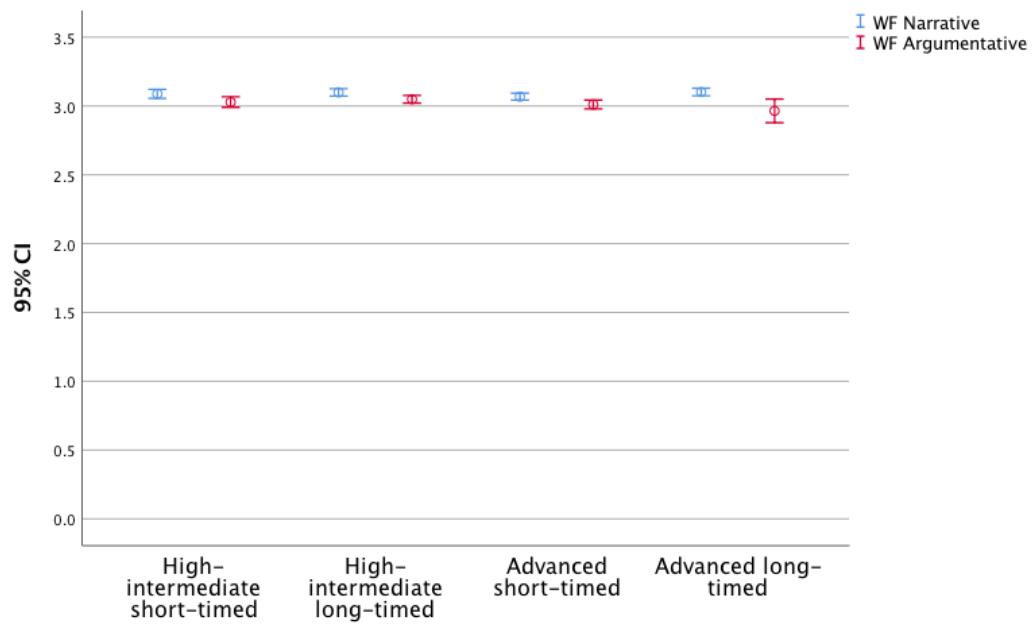
Figure 7. Genre differences in WF.
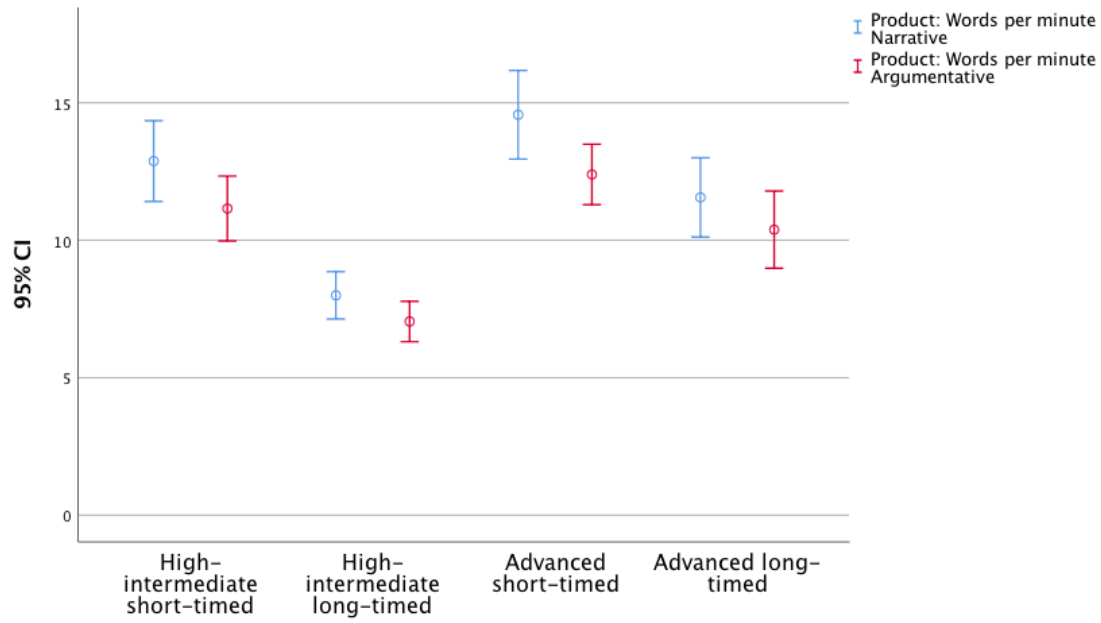
Figure 8. Genre differences in Product: Words per minute
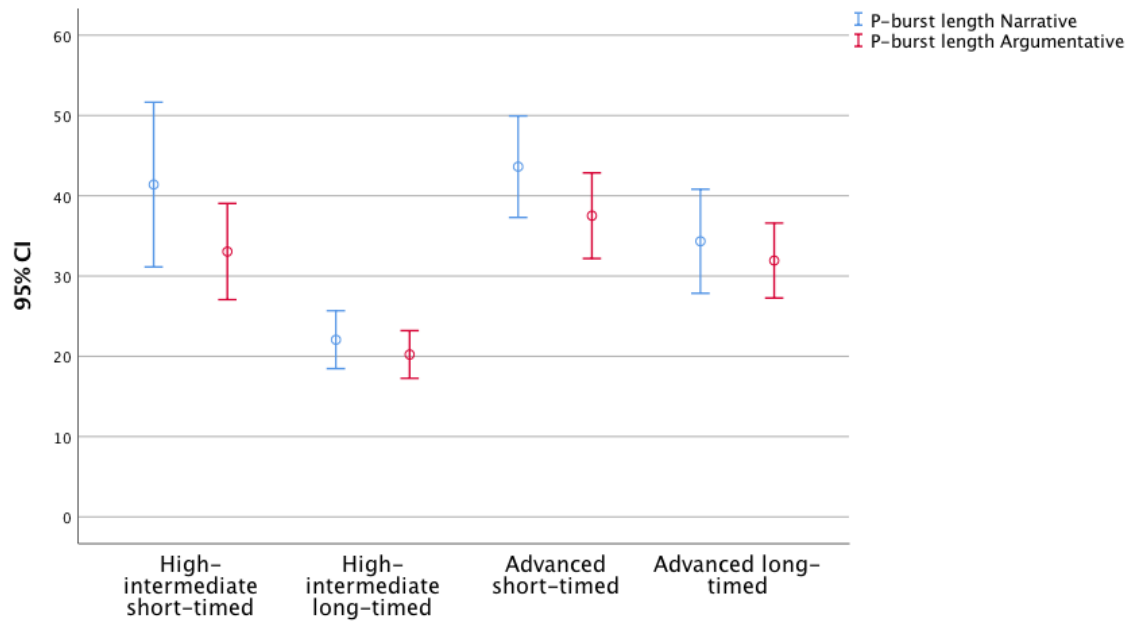
Figure 9. Genre differences in P-burst length
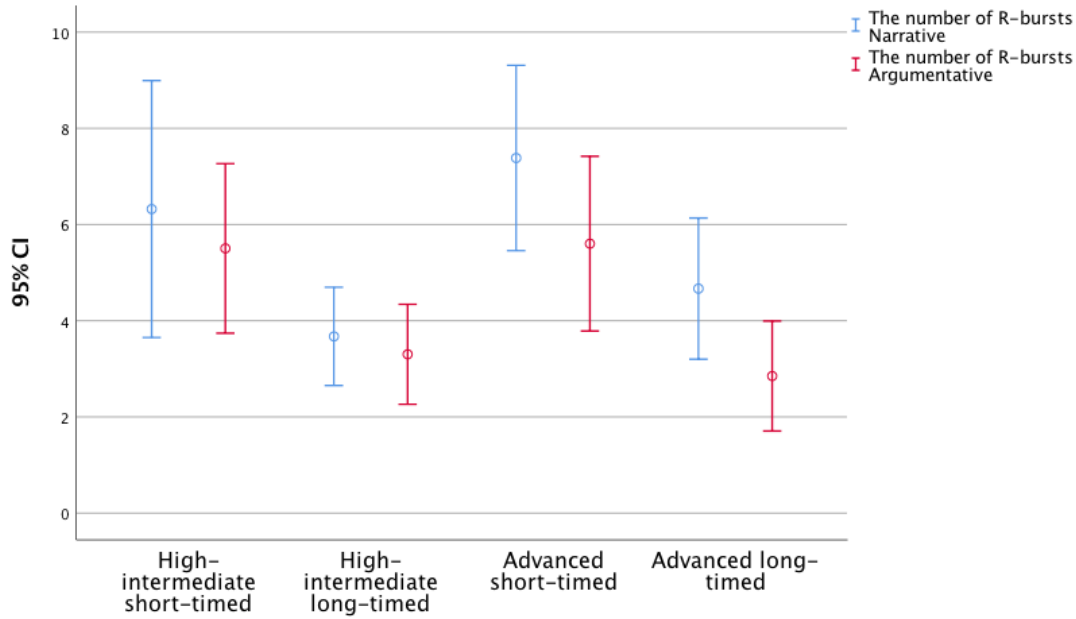
Figure 10. Genre differences in the number of R-bursts.

In order to find the effects of time constraints and proficiency on linguistic features in both genres, tests of between-subjects effects were conducted next. The MANOVA indicated statistically significant proficiency differences of the combined dependent variables according to Wilks' Lambda (.626; $F(20, 100) = 2.993$, $p = <.001$, $d = .31$). As shown in Table 12, follow-up univariate ANOVAs indicated statistically significant advantages for the advanced groups in four syntactic complexity measures and two fluency measures: MLS ($p = < .001$), MLC ($p = < .001$), CN/T ($p = < .001$), VP/T ($p = .001$), product: words per minute ($p = < .001$), and product: words per minute ($p = < .001$), but not in writing fluency behaviors (see Figures 16, 17, 18, 19, 20, and 21). Comparisons of effect sizes suggest that proficiency had a medium effect on syntactic complexity and fluency. In addition, the MANOVA indicated statistically significant time constraint differences on the combined dependent variables according to Wilks' Lambda (.576; $F(20, 100) = 3.674$, $p = < .001$, $d = .35$). The follow-up univariate ANOVAs indicated

statistically significant time constraint effects on process: words per minute ($p = <.001$), product: words per minute ($p = <.001$), P-burst length ($p = <.001$), the number of pauses between words ($p = <.001$), and the number of R-bursts ($p = <.001$). The participants in the short-timed groups showed higher fluency than those in the long-timed groups (see Figures 11 and 12). For the writing fluency behaviors, the participants in the short-timed groups paused more between words and revised more than those in the long-timed groups (see Figures 13, 14, and 15). However, the interaction between proficiency and time constraints was not statistically significant according to Wilks' Lambda (.584; $F(20, 100) = 1.209$, $p = <.001$, $d = .34$).

Table 12

MANOVA: Effects of Time Constraints and Proficiency on Linguistic Features

Measures	Proficiency			Time			Proficiency * Time		
	F	*p*	*D*	*F*	*p*	*d*	*F*	*p*	*d*
MLS	14.514	<.001*	.69	1.691	.196	.23	.182	.670	.07
MLC	15.123	<.001*	.70	1.598	.209	.23	.013	.910	.02
DC/C	7.367	.008	.49	.035	.851	.03	.093	.761	.05
CP/C	8.619	.004	.53	.608	.437	.14	3.309	.071	.33
CN/T	14.729	<.001*	.69	1.934	.167	.25	.949	.332	.18
VP/T	12.385	.001*	.64	1.414	.237	.21	.286	.593	.10
WL	5.362	.022	.42	1.715	.193	.24	.275	.601	.09
WF	3.657	.058	.35	.088	.768	.05	.480	.490	.12
D	2.241	.137	.27	.158	.692	.07	.356	.552	.11
Process: Words per minute	30.450	<.001*	1.00	33.724	<.001*	1.05	2.459	.120	.28
Product: Words per minute	18.936	<.001*	.79	38.475	<.001*	1.12	3.097	.081	.32
Number of P-bursts	1.357	.246	.21	2.243	.137	.27	1.583	.211	.23
P-burst length	8.062	.005	.51	18.971	<.001*	.79	2.556	.113	.29
Pause within words	.000	.997	0	2.050	.155	.25	.017	.897	.02
Pause between words	.730	.395	.15	20.390	<.001*	.81	1.157	.284	.19
Pause between sentences	1.961	.164	.25	3.349	.070	.33	1.864	.175	.24
Pause between paragraphs	.827	.365	.16	5.966	.016	.44	.380	.539	.11
Ratio of process and product	.024	.877	.03	5.173	.025	.41	1.501	.223	.22
Number of R-bursts	.311	.578	.10	11.433	.001*	.61	.041	.839	.04
R-burst length	1.912	.169	.14	.001	.972	0	1.751	.188	.24
Wilk's Lambda	2.589	.001*	.29	3.679	<.001*	.35	1.216	.257	.20

* *p* < .0025 (Bonferroni adjustment for dependent variables)
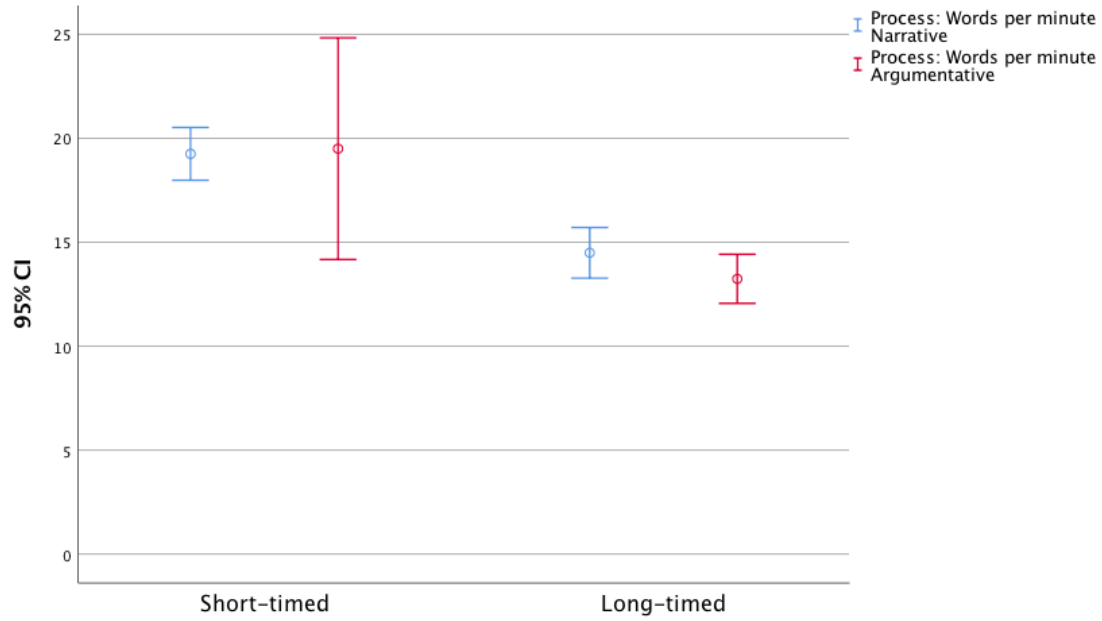
Figure 11. Effects of time constraints on process: words per minute.
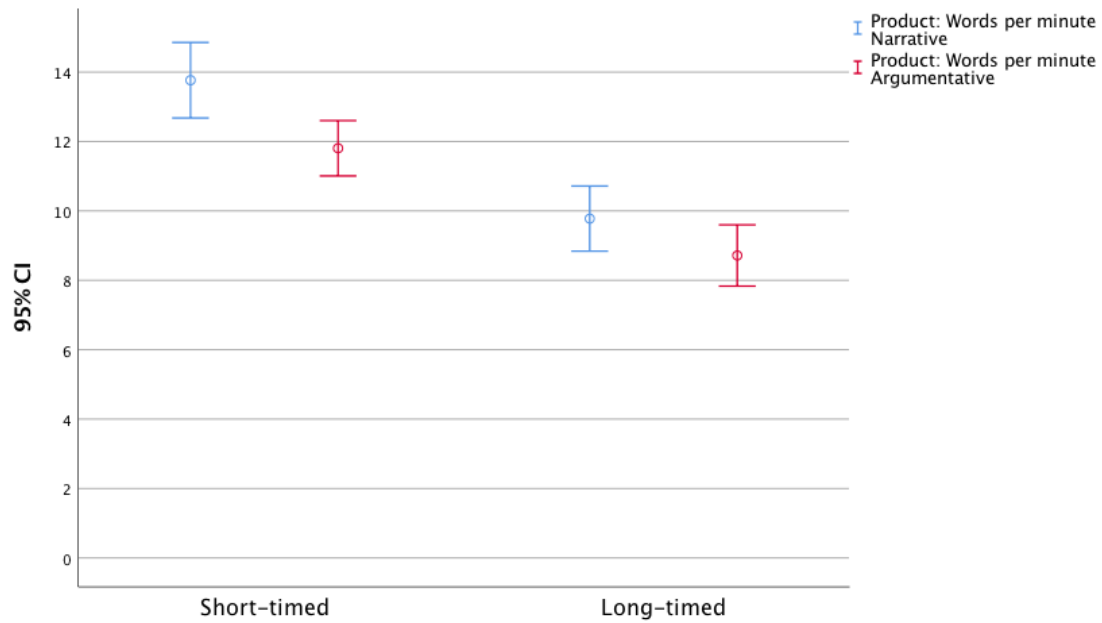
Figure 12. Effects of time constraints on product: words per minute
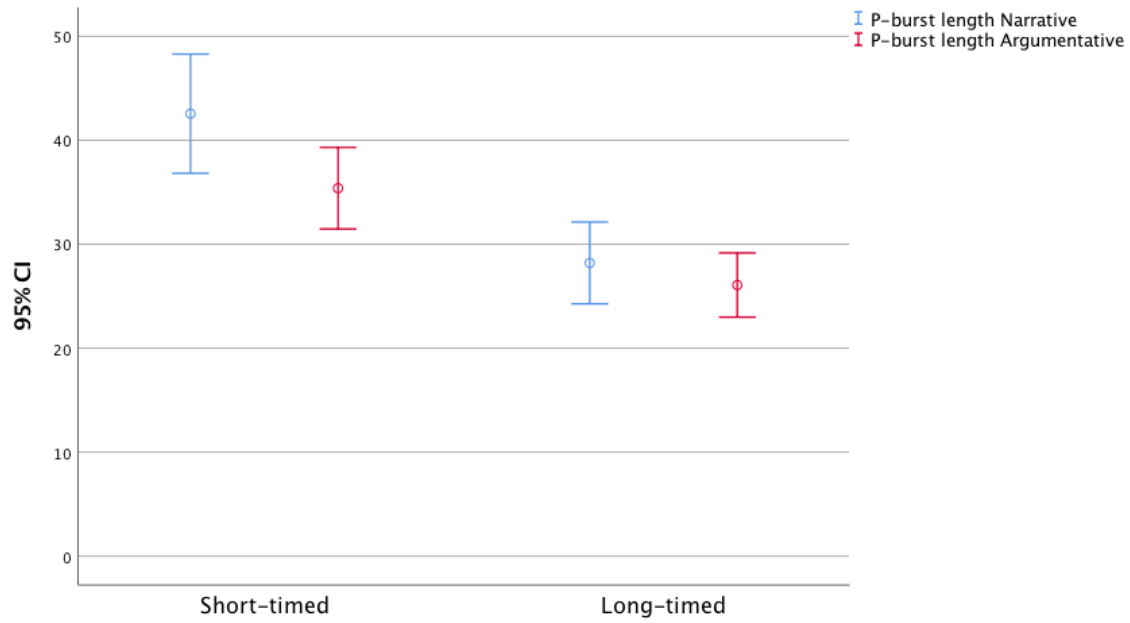
Figure 13. Effects of time constraints on p-burst length
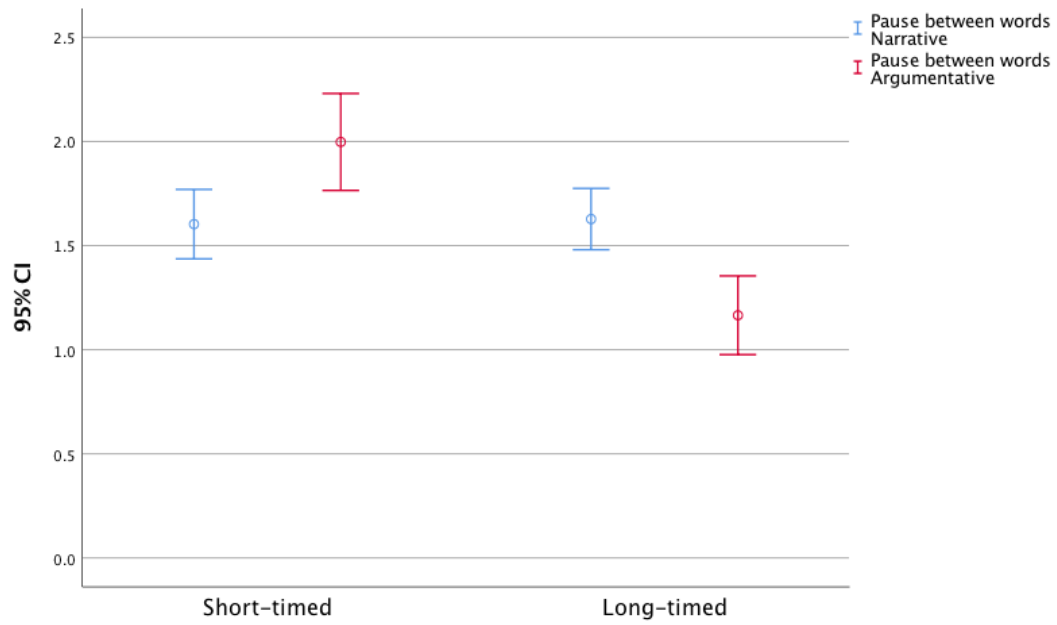
Figure 14. Effects of time constraints on pause between words
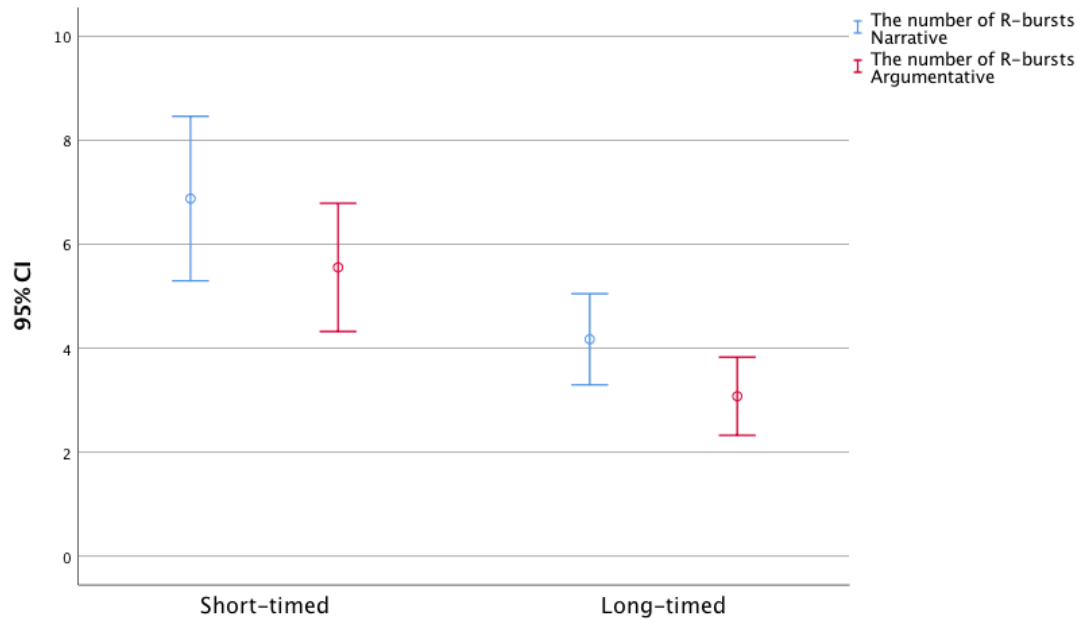
Figure 15. Effects of time constraints on the number of R-bursts.
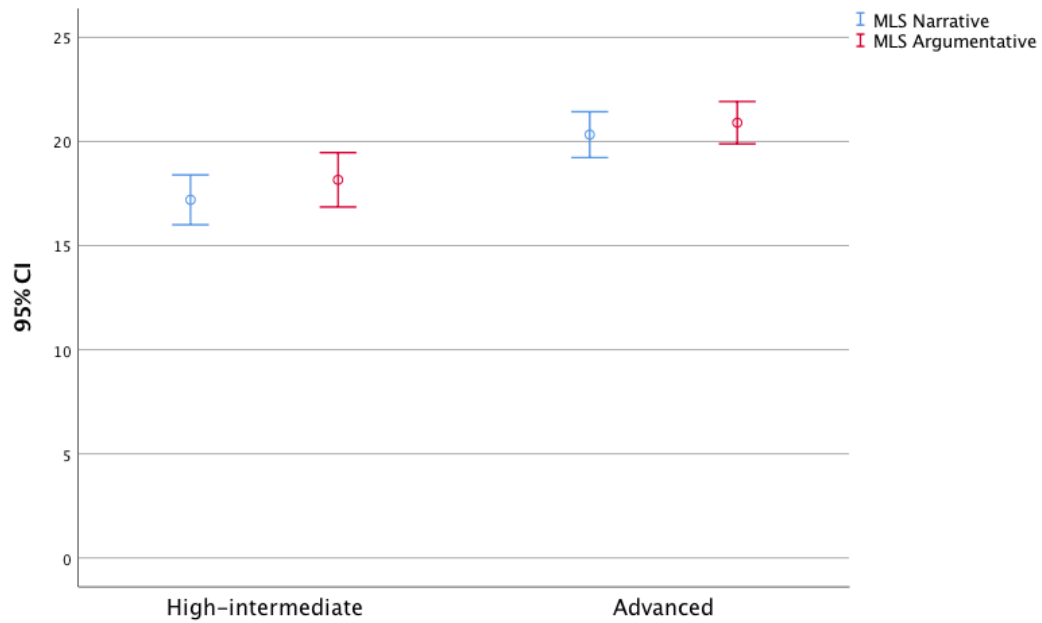
Figure 16. Effects of proficiency on MLS
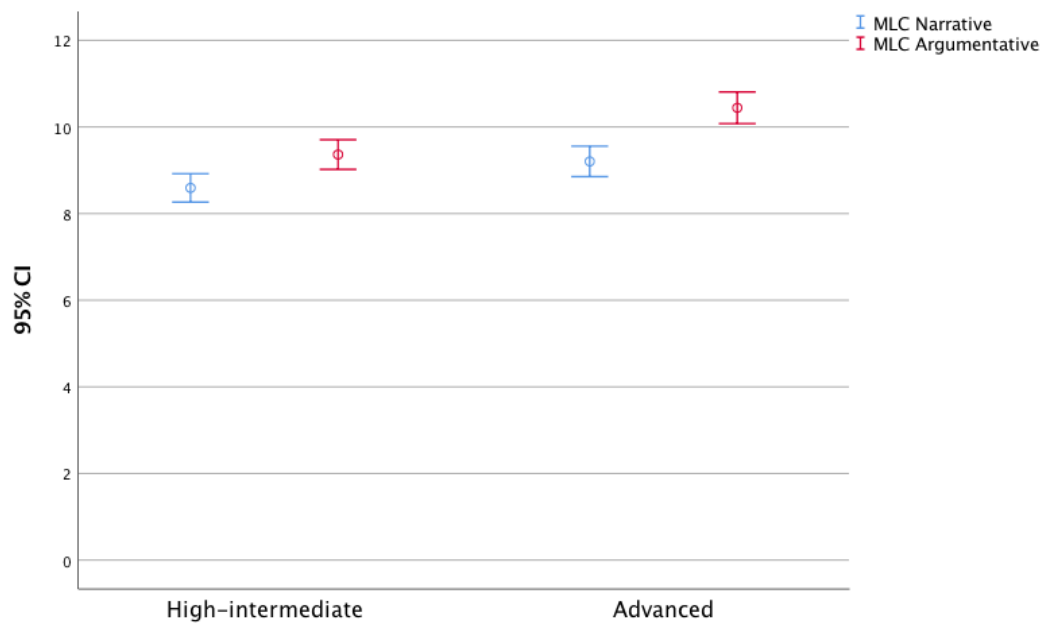
Figure 17. Effects of proficiency on MLC
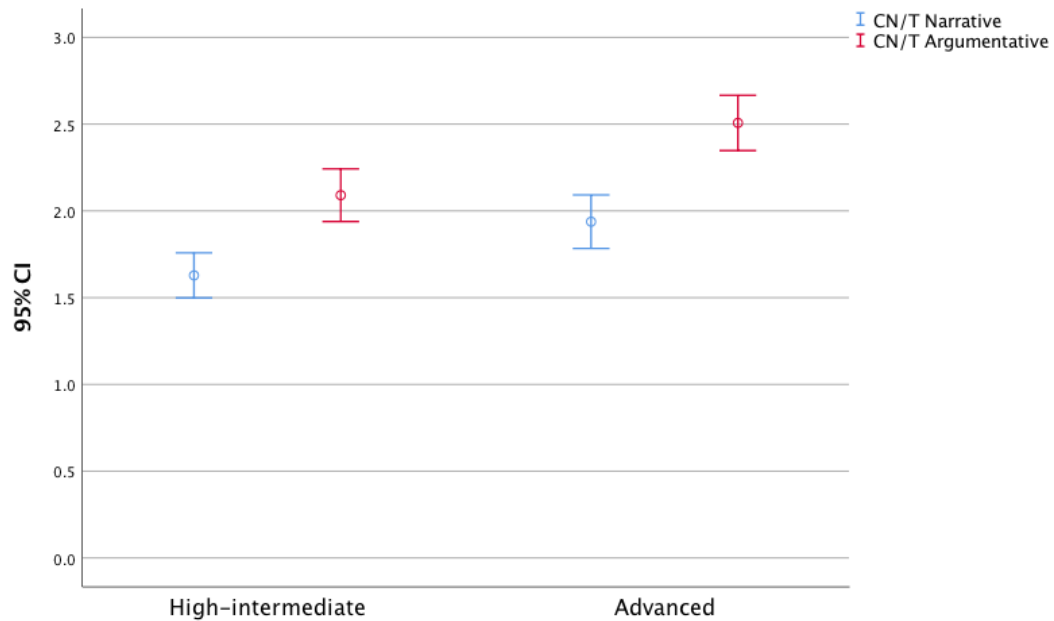
Figure 18. Effects of proficiency on CN/T
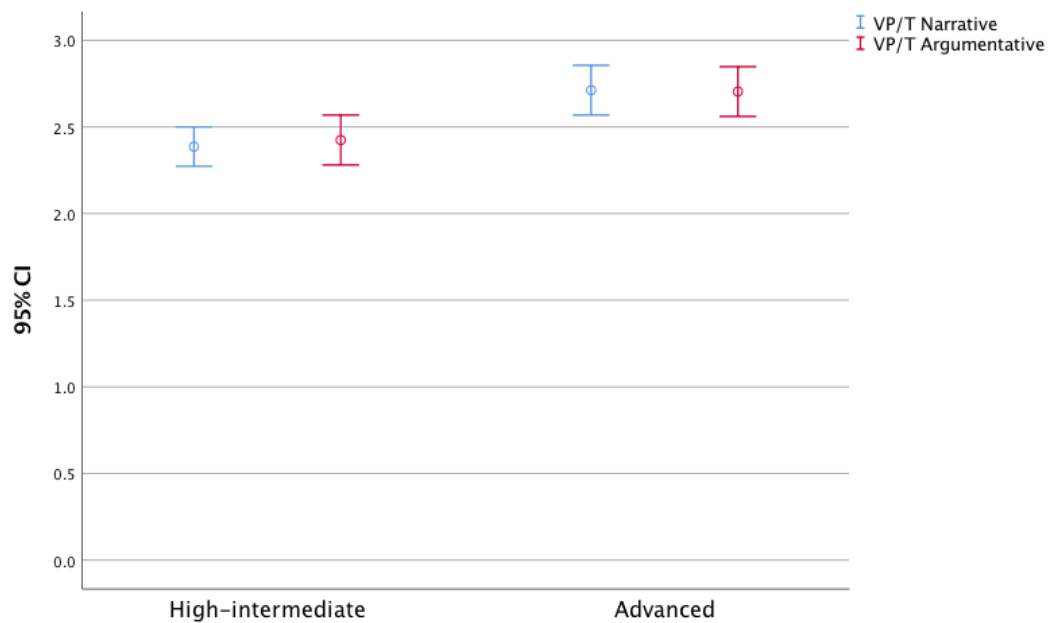
Figure 19. Effects of proficiency on VP/T
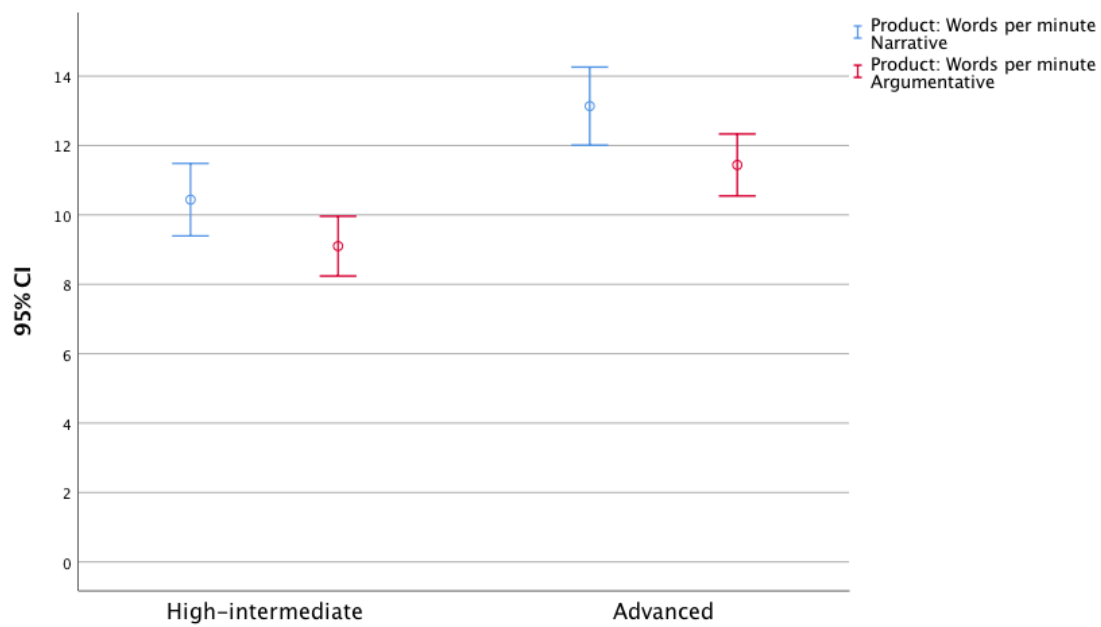
Figure 20. Effects of proficiency on product: words per minute
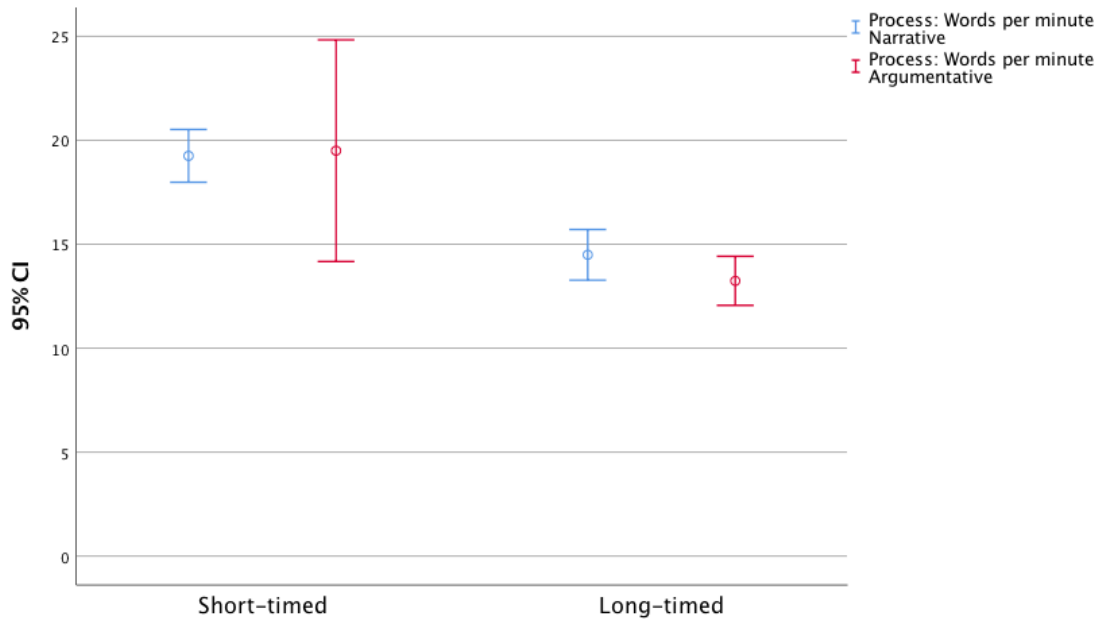
Figure 21. Effects of proficiency on process: words per minute

Table 13 presents the descriptive statistics for writing quality by time constraints, proficiency, and genres. At a glance, the advanced students received higher scores on all subscales than the high-intermediate students. The 95% confidence intervals for the total scores from high-intermediate and advanced groups do not overlap with each other, and thus there seem to be differences between the groups (see Figure 22). For the comparison between the two genres and time constraints, the mean and 95% confidence intervals did overlap across the groups.

Table 14 presents the results of the two-way ANOVA that was conducted to find effects of time constraints and proficiency on the writing quality of narratives. A main effect of proficiency was shown based on the total scores and subscale scores, with medium to large effect sizes, whereas the effect of time and the interaction of proficiency and time were not statistically significant. For the proficiency effect on narratives, advanced learners gained higher scores overall ($F(1, 119) = 35.610$, $p = < .001$, $d = 1.08$), and on content ($F(1, 119) = 27.712$, $p = < .001$,

$d = .95$), organization ($F(1, 119) = 35.073$, $p = < .001$, $d = 1.07$), vocabulary ($F(1, 119) = 28.836$, $p = < .001$, $d = .97$), language use ($F(1, 119) = 37.302$, $p = < .001$, $d = 1.10$), and mechanics ($F(1, 119) = 15.155$, $p = < .001$, $d = .70$). The total scores and all the subscale scores except for mechanics were found to have large effect sizes; mechanics had a medium effect size.

Table 13

Descriptive Statistics: Writing Quality by Time Constraints, Proficiency, and Genres

Measures	High-intermediate short-timed (N = 30)				High-intermediate long-timed (N = 30)				Advanced short-timed (N = 33)				Advanced long-timed (N = 30)			
	Nar		Arg		Nar		Arg		Nar		Arg		Nar		Arg	
	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI
Total	69.70 (6.49)	67.28, 68.62	70.13 (5.32)	68.14, 72.11	69.63 (5.74)	67.48, 71.77	70.69 (4.98)	68.83, 72.55	75.30 (6.69)	72.92, 77.67	75.31 (6.43)	73.03, 77.59	77.46 (5.90)	75.26, 79.67	78.09 (8.00)	75.10, 81.08
Content	15.98 (1.61)	15.38, 16.59	15.80 (1.39)	15.28, 16.32	15.66 (1.37)	15.15, 16.18	15.68 (1.13)	15.26, 16.10	16.97 (1.55)	16.42, 17.52	16.84 (1.60)	16.27, 17.41	17.52 (1.41)	16.99, 18.05	17.41 (1.92)	16.69, 18.12
Organization	15.37 (1.52)	14.78, 15.93	15.65 (1.30)	15.16, 16.14	15.53 (1.36)	15.02, 16.04	15.85 (1.24)	15.38, 16.31	16.67 (1.59)	16.10, 17.23	16.74 (1.61)	16.17, 17.31	17.43 (1.49)	16.88, 17.99	17.44 (1.79)	16.77, 18.11
Vocabulary	15.35 (1.37)	14.84, 15.86	15.57 (1.03)	15.18, 15.95	15.57 (1.22)	15.11, 16.02	15.62 (1.00)	15.24, 15.99	16.50 (1.46)	15.98, 17.02	16.55 (1.45)	16.03, 17.06	17.05 (1.36)	16.54, 17.56	17.28 (1.78)	16.62, 17.95
Language use	14.92 (1.40)	14.39, 15.44	14.98 (1.26)	14.51, 15.45	14.83 (1.49)	14.28, 15.39	15.28 (1.30)	14.80, 15.77	16.44 (1.63)	15.86, 17.02	16.45 (1.44)	15.94, 16.96	16.67 (1.54)	16.10, 17.24	16.97 (1.86)	16.28, 17.66
Mechanics	8.09 (1.32)	7.60, 8.58	8.13 (1.01)	7.75, 8.51	8.03 (.95)	7.68, 8.39	8.26 (.86)	7.94, 8.58	8.72 (.92)	8.40, 9.05	8.72 (.79)	8.45, 9.00	8.80 (.72)	8.53, 9.07	8.99 (.94)	8.64, 9.34
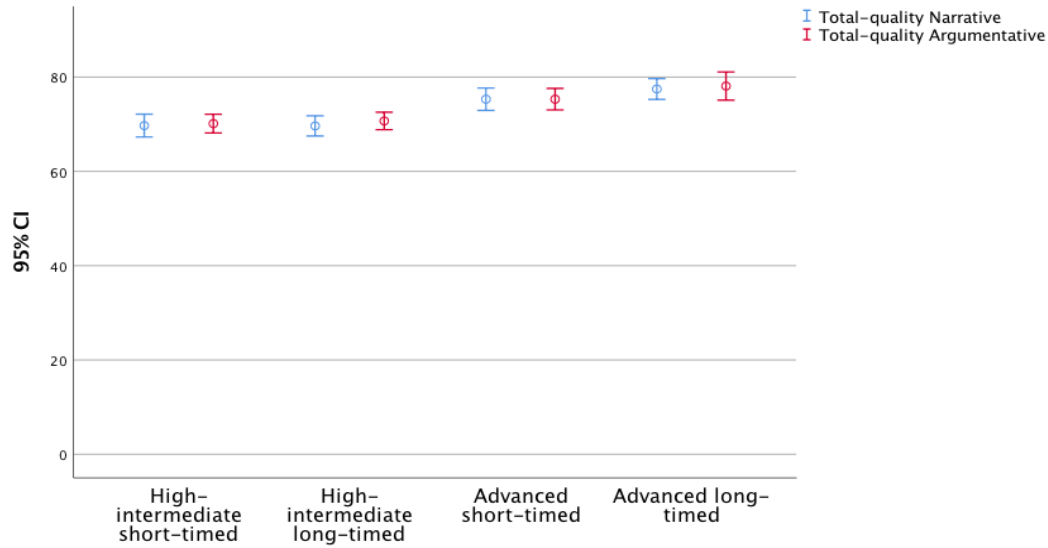
Figure 22. Total writing quality scores in the two time constraints and proficiency levels across the groups.

Table 14

Two-Way ANOVA: Effects of Time Constraints and Proficiency on Writing Quality in Narrative Essays

Measures	Proficiency			Time			Proficiency * Time		
	F	*p*	*d*	*F*	*p*	*d*	*F*	*p*	*d*
Total	35.610	<.001*	1.08	.862	.355	.17	.990	.332	.17
Content	27.712	<.001*	.95	.183	.670	.07	2.569	.112	.29
Organization	35.073	<.001*	1.07	2.984	.087	.31	1.233	.269	.20
Vocabulary	28.836	<.001*	.97	2.444	.121	.28	.462	.498	.12
Language use	37.302	<.001*	1.10	.069	.794	.05	.320	.573	.10
Mechanics	15.155	<.001*	.70	.002	.968	0	.132	.717	.07

* *p* < .0083 (Bonferroni adjustment for dependent variables)

67

Table 15

Two-Way ANOVA: Effects of Time Constraints and Proficiency on Writing Quality in

Argumentative Essays

Measures	Proficiency			Time			Proficiency * Time		
	F	*p*	*d*	*F*	*p*	*d*	*F*	*p*	*d*
Total	30.591	<.001*	.99	2.157	.145	.26	.955	.330	.18
Content	24.881	<.001*	.90	.635	.427	.14	1.480	.226	.22
Organization	24.330	<.001*	.89	2.731	.101	.30	.842	.361	.17
Vocabulary	29.153	<.001*	.97	2.586	.110	.29	1.971	.163	.25
Language use	34.667	<.001*	1.06	2.352	.128	.28	.169	.682	.07
Mechanics	16.542	<.001*	.73	1.424	.235	.22	.182	.670	.08

*$p < .0083$ (Bonferroni adjustment for dependent variables)

Table 15 presents the results of a two-way ANOVA on the effects of time constraints and proficiency on the writing quality of the argumentative essays. The results are similar to those for the narratives. A main effect of proficiency was found for the total scores and the subscale scores, with medium to large effect sizes, whereas the effect of time and the interaction of proficiency and time were not statistically significant. With regard to the proficiency effect on the argumentative essays, the advanced learners gained higher total scores ($F(1, 119) = 30.591, p = < .001, d = .99$), as well as higher scores on content ($F(1, 119) = 24.881, p = < .001, d = .90$), organization ($F(1, 119) = 24.330, p = < .001, d = .89$), vocabulary ($F(1, 119) = 29.153, p = < .001, d = .97$), language use ($F(1, 119) = 34.667, p = < .001, d = 1.06$), and mechanics ($F(1, 119) = 16.542, p = < .001, d = .73$) than the high-intermediate learners. The total scores and the subscale scores, except for mechanics, were found to have large effect sizes, while mechanics

had a medium effect size.

Table 16 presents the correlation between fluency measures and writing quality and the correlation between fluency measures and linguistic complexity in the narratives. For writing quality and fluency measures, the results of a Pearson correlation indicated significant positive associations between process: words per minute and total quality, ($r(123) = .413$, $p = < .001$), product: words per minute and total quality ($r(123) = .441$, $p = < .001$), the ratio of process and product and total quality ($r(123) = .270$, $p = .003$), P-burst length and total quality ($r(123) = .217$, $p = .016$), and R-burst length and total quality ($r(123) = .290$, $p = .001$). The two fluency measures (process and product: words per minute) showed moderate correlations with writing fluency, and the writing fluency behavior measures (the ratio of process and product, P-burst length, and R-burst length) demonstrated weak correlations. Based on Plonsky and Oswald's (2014) benchmarks for the effect size of correlation coefficients (.25: small; .40: medium; .60: large), the association between total scores for writing quality in the narratives and writing fluency behaviors had a medium or small effect. With regard to the correlation between fluency measures and linguistic complexity, there were significant associations between fluency measures (process: words per minute and product: words per minute) and syntactic complexity. In addition, writing fluency behaviors such as pausing and revision tend to be associated with lexical complexity more than with syntactic complexity. Overall, the correlation coefficients show that the association between writing fluency measures and linguistic complexity measures had a small effect size.

Table 16

Correlations: Fluency Measures with Total Writing Quality and Linguistic Complexity Measures

in Narrative Essays (N = 123)

Measures	Total quality	MLS	MLC	DC/C	CP/C	CN/T	VP/T	WL	WF	D
Process: Words per minute	.413***	.279**	.252**	.098	.191*	.190*	.249**	.158	-.120	.153
Product: Words per minute	.441***	.228*	.227*	.095	.139	.152	.205*	.089	-.107	.115
Ratio of process and product	.270**	.042	.038	.111	-.055	.051	.049	-.058	-.057	-.083
Number of P-bursts	-.063	.001	-.070	.085	-.127	.023	-.029	-.048	.060	-.159
P-burst length	.217*	.112	.187*	-.048	.154	.068	.110	.209*	-.205*	.159
Pause within words	-.118	.040	-.062	.039	.019	.058	-.032	.111	.031	-.011
Pause between words	.002	.006	-.032	.084	-.123	.020	-.008	-.086	-.001	-.218*
Pause between sentences	.019	-.188*	-.171	-.072	-.212*	-.113	-.132	-.131	-.054	-.277*
Pause between paragraphs	.046	.043	.105	.077	-.066	.178*	.135	.084	-.171	.100
Number of R-bursts	-.045	.074	.003	.036	.016	.040	.025	.141	-.180*	.169
R-burst length	.290**	-.036	.106	-.043	.141	-.042	.036	.058	-.118	.046

*** $p < 0.001$, ** $p < 0.01$, * $p < .05$

To further explore the predictive relationship between writing quality and fluency measures in narratives, a multiple regression was performed to find which fluency measures most strongly predicted the narratives' overall writing quality. A stepwise multiple regression (probability of F to enter = .05), beginning with all eleven fluency measures, identified two statistically significant models for predicting the overall writing quality of the narratives. As shown in Table 17, product: words per minute alone predicted a relatively large proportion of the variance in writing quality ($R^2 = .195$, $F(1, 121) = 29.251$, $p = < .001$). The addition of R-burst length increased the predictive power slightly ($R^2 = .228$, $F(1, 120) = 5.147$, $p = .025$). Unstandardized beta values (Table 18) indicated that an increase of one word in the R-burst

length and product: words per minute was related to an increase of between .25 and .62 score points on the writing quality of the narrative essays. The remaining nine variables did not contribute additional unique statistically significant variance once the two main predictor variables were removed from the model.

Table 17

Model Summary: Total Quality as Criterion Variable in Narrative Essays

Model	R	R^2	Adjusted R^2	Std. Error of the estimate	Change statistics				
					R^2 change	F change	$df1$	$df2$	Sig. F change
1	.441[a]	.195	.188	6.351	.195	29.251	1	121	.000
2	.477[b]	.228	.215	6.245	.033	5.147	1	120	.025

[a] Predictors: (constant), product: words per minute

[b] Predictors: (constant), product: words per minute, R-burst length

Table 18

Coefficients[a]: Total Quality as Criterion Variable in Narrative Essays

Model	Unstandardized coefficients		Standardized coefficients	t	Sig.	95% confidence interval for B		Correlations			Collinearity statistics	
	B	Std. Error	Beta			Lower bound	Upper bound	Zero-order	Partial	Part	Tolerance	VIF
1 (Constant)	64.834	1.629		39.794	<.001	61.609	68.060					
Product: Words per minute	.698	.129	.441	5.408	<.001	.442	.953	.441	.441	.441	1.000	1.000
2 (Constant)	62.772	1.842		34.080	<.001	59.125	66.419					
Product: Words per minute	.621	.131	.392	4.726	<.001	.361	.881	.441	.396	.379	.933	1.072
R-burst length	.249	.110	.188	2.269	.025	.032	.466	.290	.203	.182	.993	1.072

[a] Dependent variable: total quality

Table 19 presents the correlations between fluency measures and writing quality and the correlation between fluency measures and linguistic complexity in argumentative essays. Different from the correlation between fluency measures and total quality in narratives, the results of a Pearson correlation indicated only three significant positive associations: process: words per minute and total quality ($r(123) = .457$, $p = < .001$), product: words per minute and total quality ($r(123) = .410$, $p = < .001$) and P-burst length and total quality ($r(123) = .361$, $p < .001$). The fluency measures (process: words per minute and product: words per minute) and the writing fluency behavior measure (P-bursts) showed moderate correlations with the writing quality of the argumentative essays. With regard to the effect sizes, the association between total scores on writing quality in the argumentative essays and writing fluency had a medium effect. For the correlations between writing fluency measures and linguistic complexity, there are several significant associations: process: words per minute and MLS, product: words per minute and MLS, number of P-bursts and D, pauses within words and CP/C, and pauses between paragraphs and CN/T. The correlation coefficients showed that these associations between writing fluency measures and linguistic complexity measures had a small effect size.

Table 19

Correlations: Fluency Measures with Total Writing Quality and Linguistic Complexity Measures in Argumentative Essays (N = 123)

Measures	Total quality	MLS	MLC	DC/C	CP/C	CN/T	VP/T	WL	WF	D
Process: Words per minute	.457***	.308**	-.022	.047	-.143	.029	.057	-.100	.048	-.111
Product: Words per minute	.410***	.283**	.125	.159	.027	.147	.172	-.010	.063	.033
Ratio of process and product	.010	.049	-.107	.105	-.058	-.039	.062	-.166	-.099	-.090
Number of P-bursts	-.115	.045	-.011	.106	.069	.062	.176	-.045	-.015	-.210*
P-burst length	.361***	.194*	.141	.061	.020	.132	.050	.113	.019	.171
Pause within words	.042	-.026	-.115	.015	-.198*	-.055	-.044	-.123	.146	-.138
Pause between words	.068	.038	.088	-.008	-.055	.388	.474	.095	.830	.614
Pause between sentences	-.064	-.030	-.058	.089	-.165	.010	.011	.045	.004	.056
Pause between paragraphs	-.095	.124	-.014	.164	-.123	.197*	.121	-.051	.067	.009
Number of R-bursts	-.055	.069	.111	-.084	.010	-.008	-.011	.033	.060	-.003
R-burst length	.169	.154	-.062	.210*	.105	.068	.185*	-.043	-.065	-.064

*** $p < 0.001$, ** $p < 0.01$, * $p < .05$

In order to investigate the predictive relationship between writing quality and fluency measures for argumentative essays, a multiple regression was performed to find which fluency measures most strongly predicted the argumentative essays' overall writing quality. A stepwise multiple regression (probability of F to enter = .05), beginning with all eleven fluency measures,

identified two statistically significant models for predicting the overall writing quality of the

argumentative essays. As shown in Table 20, process: words per minute alone predicted a

relatively large proportion of the variance in writing quality of the argumentative essays (R^2

$= .209$, $F(1, 121) = 31.991$, $p = < .001$). The addition of the number of R-bursts ($R^2 = .260$, $F(1,$

$120) = 8.240$, $p = .005$) increased the predictive power slightly. Unstandardized beta values

(Table 21) indicated that an increase of one word in process: words per minute and a decrease of

one point of the number of R-bursts were related to a change of between -.400 and .790 score

points on the writing quality of the argumentative essays. The remaining nine variables did not

contribute additional unique statistically significant variance once the two main predictor

variables were removed from the model.

Table 20

Model Summary: Total Quality as Criterion Variable in Argumentative Essays

Model	R	R^2	Adjusted R^2	Std. Error of the estimate	Change statistics				
					R^2 Change	F change	$df1$	$df2$	Sig. F change
1	.457[a]	.209	.203	6.282	.209	31.991	1	121	.000
2	.510[b]	.260	.248	6.102	.051	8.240	1	120	.005

[a] Predictors: (constant), process: words per minute
[b] Predictors: (constant), process: words per minute, the number of R-bursts

Table 21

Coefficients[a]: Total Quality as Criterion Variable in Argumentative Essays

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.	95% confidence interval for *B*		Correlations			Collinearity statistics	
		B	Std. Error	Beta			Lower bound	Upper bound	Zero-order	Partial	Part	Toler-ance	VIF
1	(Constant)	63.485	1.876		33.844	<.001	59.772	67.199					
	Process: Words per minute	.669	.118	.457	5.656	<.001	.435	.903	.457	.457	.457	1.000	1.000
2	(Constant)	63.400	1.822		34.790	<.001	59.791	67.008					
	Process: Words per minute	.790	.122	.540	6.455	<.001	.547	1.032	.457	.508	.507	.882	1.134
	The number of R-bursts	−.400	.139	−.240	−2.871	.005	−.676	−.124	−.005	−.253	−.225	.882	1.134

[a] Dependent variable: total quality

76

4.2. Qualitative analysis

The stimulated recall data ($N = 16$) were used to triangulate the quantitative results regarding the second research question. Based on Kellogg's (1996) model of writing, the participants' comments on pausing and revision were categorized as pertaining to planning, translation, or monitoring processes. Within the planning processes (planning and organization), the majority of recall comments were about content, and within the translation processes (lexical retrieval, syntactic encoding, and cohesion), more than half of the comments were about lexical retrieval.

Table 22 presents examples from Participant #7's argumentative essay, classified by types of writing process, along with the participant's stimulated recall comments regarding pausing while writing these examples. The first example is from the first paragraph; the participant said he had decided to argue against the prompt, and he paused between words to plan what content and supporting ideas he would use to disagree with the prompt's statement regarding the necessity of foreign language abilities. The next example shows a translation process; here, he paused to search for synonyms for "foreign language," because he did not want to use the same words repeatedly; however, he did not find an appropriate synonym. The third example illustrates pausing for the purpose of monitoring; as the participant's stimulated recall shows, he paused because he noted an error; he thought the term "world trade sector" was not appropriate. He later changed "world trade sector" to "trade sector."

Table 22

Pausing: Writing Processes, Text Examples, and Stimulated Recall Comments (Participant #7)

Writing process	Text	Stimulated recall comments
Planning	I could bet that a lot of students would agree that foreign language abilities are necessary in this globalized area. However, is it that much? Many academies in…(pause)	I decided a position to write. I came up with contents and supporting ideas. I wanted to oppose the prompt.
Translation	However, is it that much? This essay would talk about it is not that necessary… (pause) Many English academies in Korea wants to	I tried to think of vocabulary that substitutes "foreign language", but it was hard to find one. I wanted to write a different word that means foreign language, but it was hard to find one. I would write the same vocabulary, "to use foreign language".
Monitoring	World trade sector in world economy structure becomes larger…(pause)	As I read, I found an error and wanted to go back and fix it.

Table 23 presents examples of the writing processes related to revision, and the stimulated recall comments regarding these specific revisions in the argumentative essay of

Participant #4. In the first example, in a process of planning while writing, the participant wrote a sentence beginning with the connective "for instance"; she then decided she should emphasize a general point before writing about specific advantages. She therefore deleted the connective and inserted a new sentence between the two sentences she had just written. Next, the participant engaged in a translation process to retrieve lexical items as she wrote. She decided to revise "person" (a singular noun) to "a group of…people" (a collective noun), because, she explained, "group" was more appropriate in the context.

Because the number of recalls was different from the number of participants (e.g., 30 minutes and 60 minutes), the data were converted to percentages. Figures 23 and 24 show the percentage of stimulated recall comments that tap into the writers' cognitive processes underlying pausing and revision behaviors in the writing of narratives and argumentative essays (the actual numbers of comments are included in the appendix I).

Figure 23 summarizes the distribution of the comments about pausing that the four groups of participants made. These stimulated recall data demonstrate that there are differences in the processes underlying pausing behaviors when participants at different proficiencies write in different genres under different time constraints. The proficiency comparisons show the advanced learners' comments about pausing are more associated with translation than the high-intermediate learners', and this is the case for both narrative and argumentative essays (narrative: 37% for advanced short-timed and 49% for advanced long-timed; argumentative: 35% for advanced short-timed and 40% for advanced long-timed). In other words, the advanced learners recalled pausing for lexical retrieval, syntactic encoding, and cohesion much more often than did the high-intermediate learners.

Table 23

Revision: Writing Processes, Text Examples, and Stimulated Recall Comments (Participant #4)

Writing process	Text	Stimulated recall comments
Planning	Especially in the current globalized era, being able to speak another language can bring much more benefits such as speaking different people around the world, visiting to other countries, and learning more about another country's culture. [deleted: *for instance*] [inserted: *In various ways, being able to speak a foreign language fluently can lead to a lot of benefits that another abilities can fulfill.*] The more language one can speak and understand, the more people that person can communicate with and learn about another language.	I wanted to put an emphasis on the need to learn a foreign language before presenting advantages of learning a foreign language. I added one sentence here.
Translation	For example, I realized the necessity of a foreign language when I met [deleted: *a Chinese person*] *a group of Chinese people* in the streets.	I deleted "person" and changed it to group because group was more appropriate in this context.

In regard to time constraints, in writing an argumentative essay, the short-timed group students' comments are more associated with planning (53% for intermediate short-timed writing and 52% for advanced short-timed writing) than the long-timed group students'. On the other hand, the long-timed group students made more comments associated with monitoring, compared to the short-timed group students.

With regard to genre differences, the distribution of translation-related pausing is similar across the groups; however, the distribution of planning-related pausing is different across the groups. During pauses, the short-timed groups showed more planning in argumentative than in narrative writing, while the long-timed groups showed more planning in narrative than in argumentative writing. In addition, unlike the short-timed groups, the long-timed groups showed more monitoring-related pausing in argumentative essays than in narrative essays.
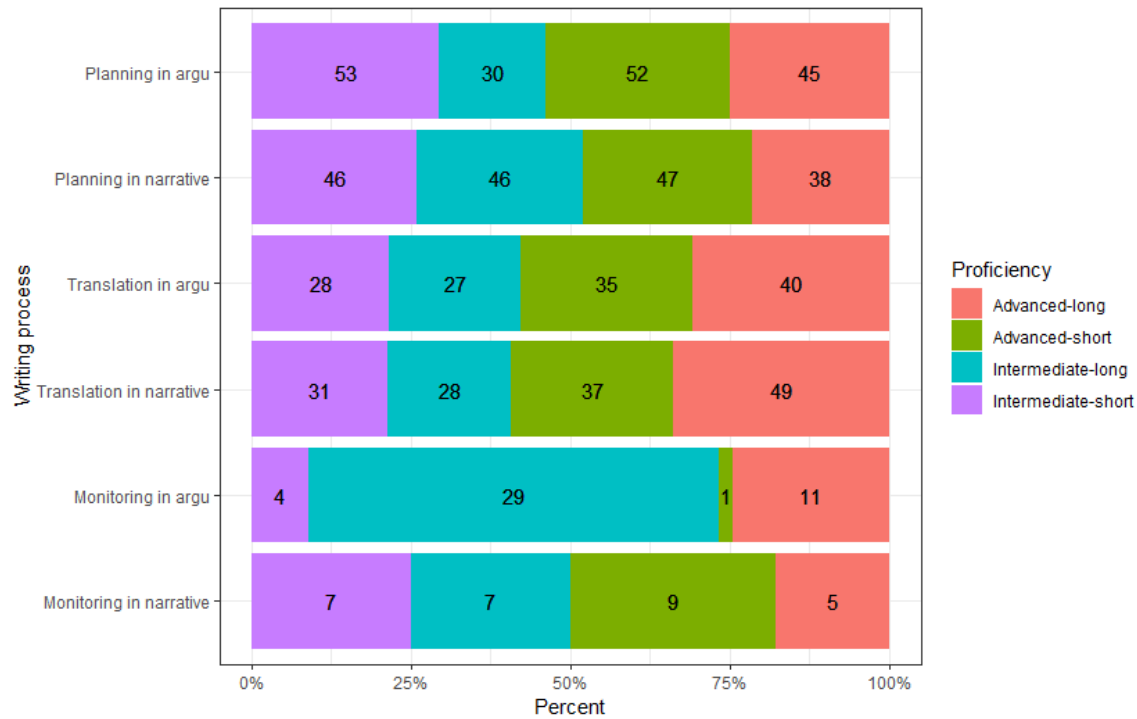
Figure 23. Comments about pausing from stimulated-recall sessions.

Figure 24 shows the distribution of comments about revision from the stimulated recall sessions. These comments suggest similarities and differences in the processes underlying revision behaviors when participants at different proficiency levels were writing in different genres under different time constraints.
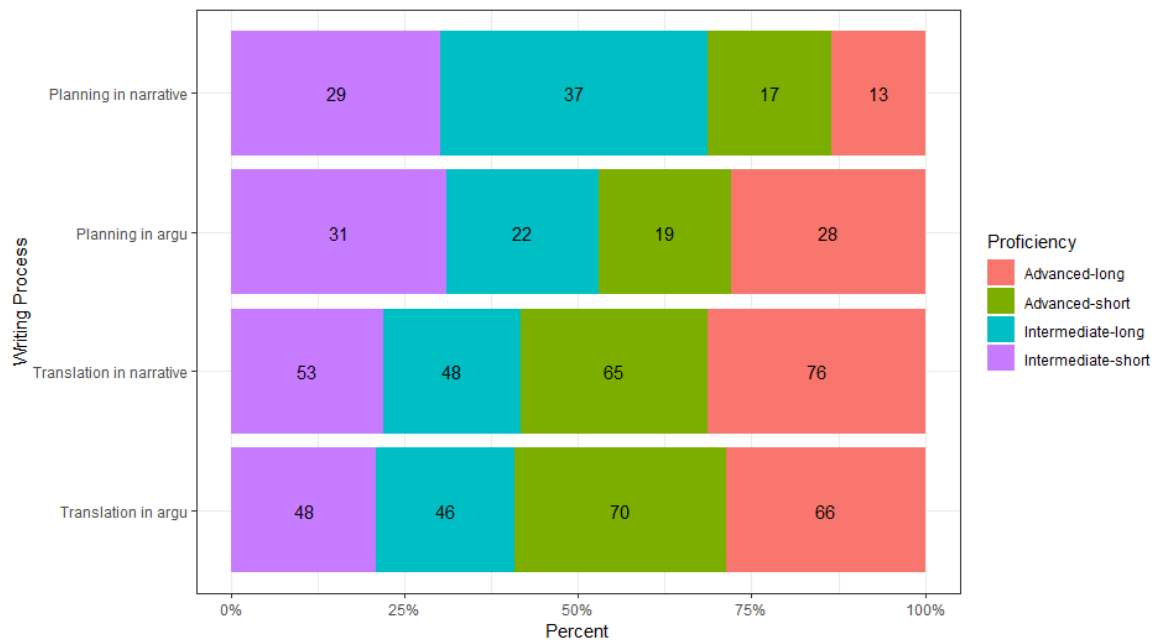
Figure 24. Comments about revision from stimulated-recall sessions.

Overall, in contrast to the comments about pausing, a higher percentage of comments regarding revision referred to translation than to planning across all groups. Compared to the number of revision comments on planning processes, the participants made more comments about translation processes (lexical retrieval, syntactic encoding, and cohesion). In particular, it is worth focusing on proficiency differences in planning and translation: The advanced students tended to make more comments on translation than the high-intermediate students in both genres.

Regarding time constraint differences, the long-timed groups and the short-timed groups did not show much difference in translation processes. For planning, there are some differences

depending on proficiency. The intermediate long-timed and short-timed groups showed similar amounts of translation-related revision comments; however, in writing narratives, the long-timed learner groups showed more planning than the short-timed learner groups, but the long-timed groups showed less planning in argumentative writing. On the other hand, the advanced long-timed group made more comments on planning than the advanced short-timed group during the writing of argumentative essays. However, in contrast to the advanced short-timed group students, the advanced long-timed group students made fewer comments about planning and more comments about translation in relation to revision when writing their narratives.

With regard to genre differences in recall comments about revision, the difference is not large in numbers of comments related to the translation process; however, there is some difference in planning processes. The learners in the high-intermediate long-timed group commented more on planning in narratives (37%) than in argumentative essays (22%). However, those in the advanced long-timed group made more comments related to planning for the argumentative essays (28%) than for the narratives (13%).

The participants' comments about pausing and revision showed patterns according to the locations of pauses and revision behaviors (Figures 25, 26, 27, and 28). Kellogg's (1996) model of writing describes cognitive processes during writing as lower or higher processes. In the stimulated recall data, higher level textual units such as sentences are associated with the comments regarding higher levels of writing processes such as planning rather than translation, regardless of time constraints, genres, and proficiencies.

Regarding their pausing, most of the participants' comments about translation and about planning occurred in different textual locations. The participants made more comments related to translation to explain pauses between words in both genres, and most of their comments related

to translation processes such as lexical retrieval, syntactic encoding, and cohesion referred to pauses they made between words. In regard to translation processes, few comments were made between sentences. Many of their comments associated with planning were made to explain their pauses between words and between clauses, although some referred to pauses between sentences.
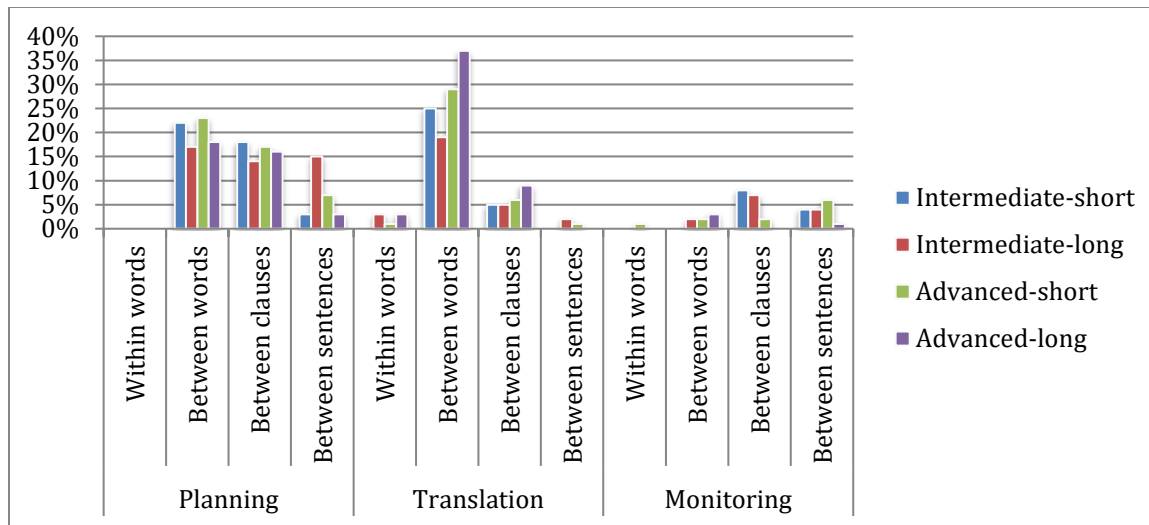
Figure 25. Comments about pausing in narratives.
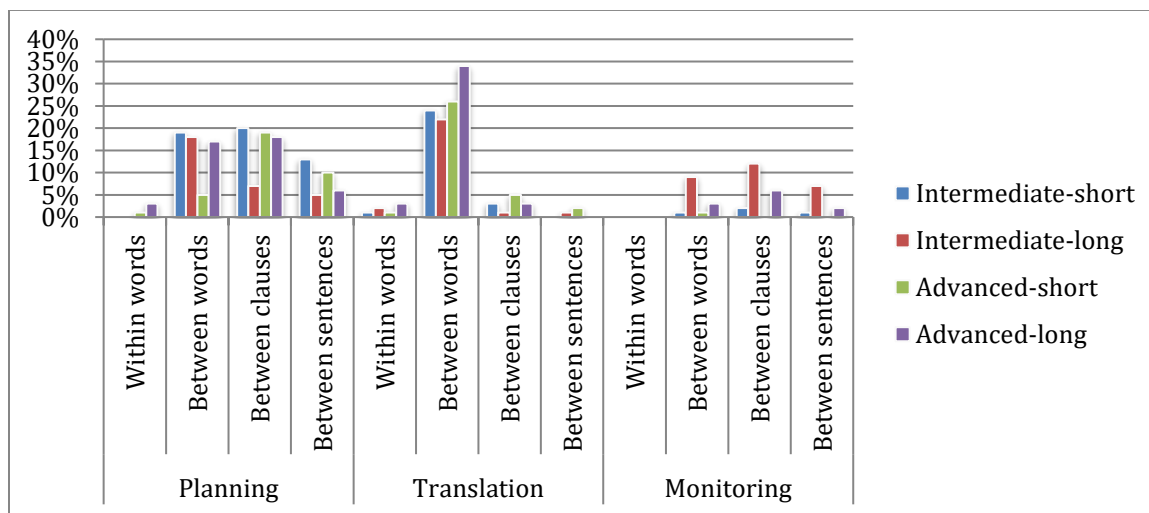
Figure 26. Comments about pausing in argumentative essays.

With regard to how comments about revision aligned with textual locations, the participants showed different patterns in the narratives and the argumentative essays, though all groups spent more time on translation than planning process. In narratives, the high-intermediate long-timed group participants made more comments about planning at the word and sentence levels than those in the other three groups. The long-timed groups' comments showed more translation below the word level than the comments of the short-timed groups. Comparing the proficiency levels, the advanced groups made more comments on translation at the word level and below the sentence level than did the high-intermediate groups. Advanced long-timed group participants commented more about revision at the clause and sentence levels compared to the other three groups. The advanced long-timed group made more comments about planning below the clause level than the other groups. In the argumentative essays, the two long-timed groups made more comments about translation at the word and below the clause level than the two short-timed groups.
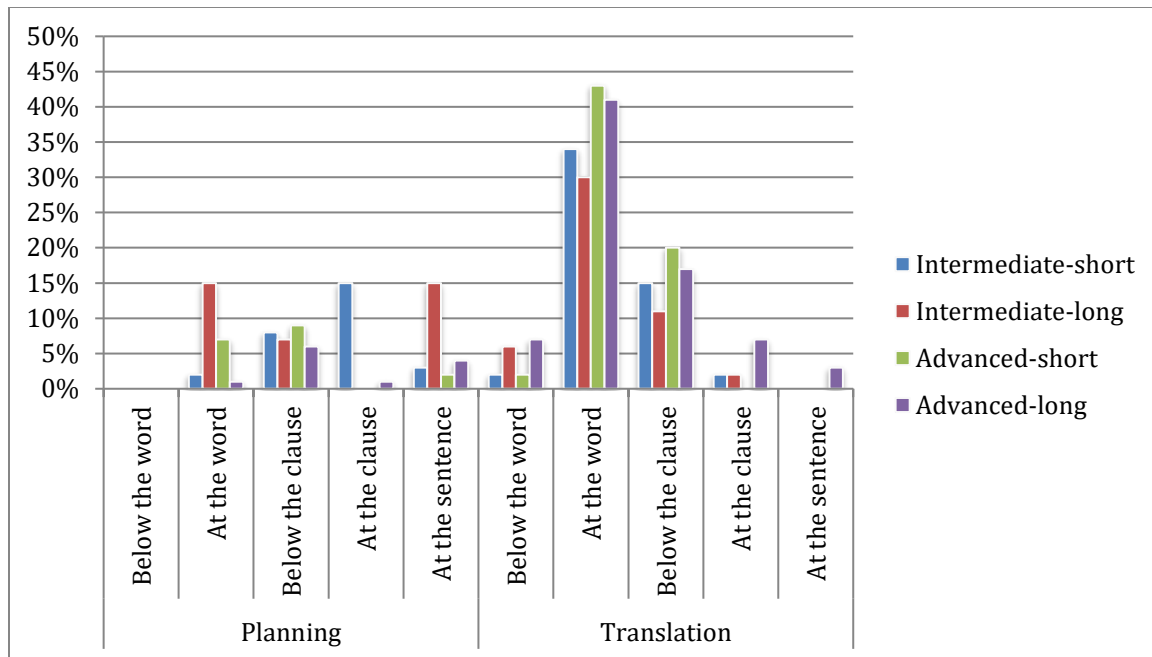
Figure 27. Comments about revision in narratives.
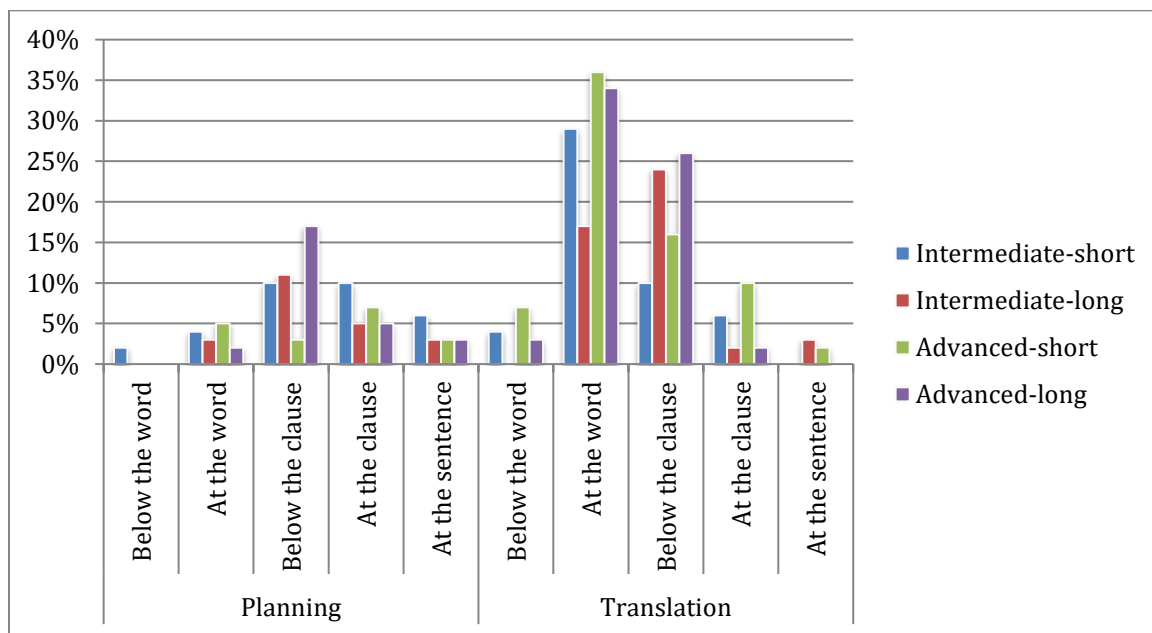
Figure 28. Comments about revision in argumentative essays.

Because the MANOVA showed a significant interaction between time constraints and genre, the qualitative data were used to learn more about how the participants used the time during pauses between words (see Figure 29). Because the number of recalls was different in the two time constraint groups, data were converted to percentages. At a glance, all learners showed similar patterns for planning, translation, and monitoring across the two genres. All groups spent more time on translation for both genres and spent less time on planning and monitoring for both genres. Although each group included only two students, the advanced learners showed more similar patterns during pauses between words than the high-intermediate learners when they were writing in the two different genres.
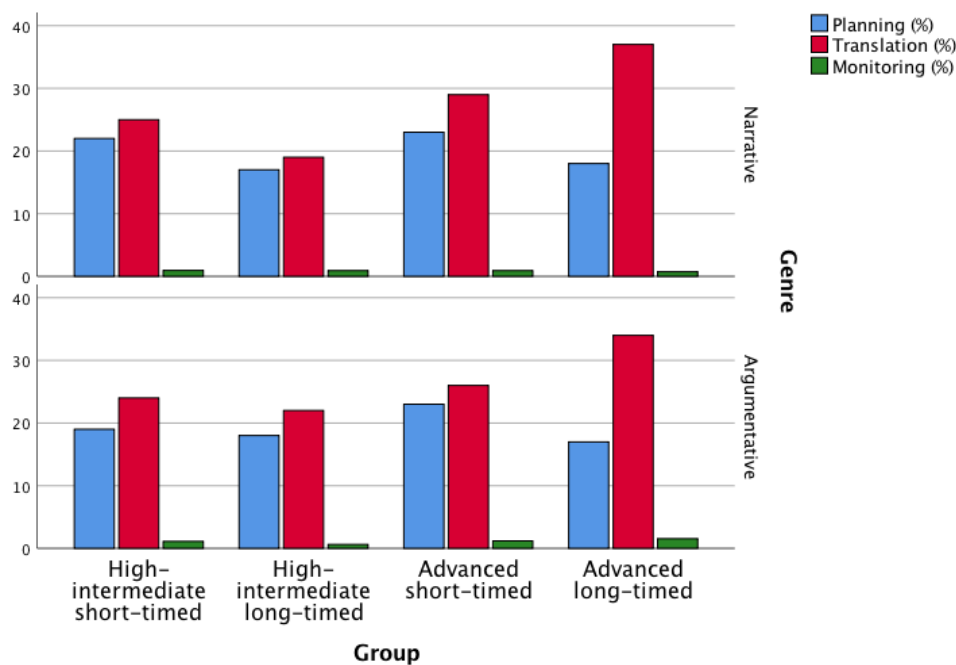
Figure 29. Writing processes during pauses between words.

4.3. Exit questionnaire results: L2 writers' perceptions of the time constraints and genres

The exit questionnaire collected data on the participants' perceptions of the genres and time constraints. Tables 24 and 25 reflect the results from the two open-ended questions, and Table 26 shows the descriptive statistics of the responses to the eight Likert-scale items. To analyze the Likert-scale data, one-way analyses of variance (ANOVA) and Bonferroni post-hoc tests were used to look for differences in the learners' perceptions.

Table 24

Questionnaire Responses by Group: "How did you feel about writing narrative and argumentative essays? Is one type of essay writing more difficult than the other?"

Group	Narrative is more difficult	Argumentative is more difficult	Both are similarly difficult
High-intermediate short-timed (*N* = 30)	20%	77%	3%
High-intermediate long-timed (*N* = 30)	27%	60%	13%
Advanced short-timed (*N* = 33)	40%	57%	3%
Advanced long-timed (*N* = 30)	40%	53%	7%

With regard to the perceived difficulty of writing in the two genres, more than half of the participants considered the argumentative genre more difficult than the narrative genre. Although this perception seems to have varied depending on the participants' English proficiency, many high-intermediate and advanced students tended to feel that argumentative essays were more

difficult to write, as illustrated in Excerpts 1 and 2.

Excerpt 1. High-intermediate student in short-timed group, #119

"In writing the narrative, it was possible to write naturally as my brainstorming process connected to my writing process smoothly. But it was difficult to write and revise the argumentative essay when I brainstormed ideas and thought about the logical flow of writing."

Excerpt 2. Advanced student in short-timed group, #16

"It was difficult for me to write the argumentative essay. In writing the argumentative essay, I needed to think about language expressions appropriate for academic writing. However, in writing a narrative, I was able to use colloquial expressions as I talk to my friends. Writing the narrative was easier than writing the argumentative essay."

As shown in Table 24, more than half of the participants in the long-timed group considered the time allotment enough for both genres. However, the participants in the short-timed group felt that the time allowed was only enough for the narrative essay. In other words, they wanted more time for writing the argumentative essay. In the next two excerpts, an intermediate learner (Excerpt 3), and an advanced learner (Excerpt 4) explain why they wanted more time for writing their argumentative essays. The participants' responses show that the learners in the short-timed group were aware of genre differences, as they wanted time for different kinds of writing processes, such as selecting vocabulary, to meet demands specific to the argumentative genre.

Table 25

Questionnaire Responses by Group: "Do you think the time allotted was enough to write the essays (both genres)?"

Group	Enough for both genres	Enough for the narrative essay only	Enough for the argumentative essay only	Not enough for both genres
High-intermediate short-timed (*N* = 30)	33%	27%	10%	30%
High-intermediate long-timed (*N* = 30)	50%	33%	10%	7%
Advanced short-timed (*N* = 33)	36%	30%	12%	21%
Advanced long-timed (*N* = 30)	73%	17%	7%	3%

Excerpt 3. Intermediate student in long-timed group, #7

"After writing the narrative, I had some spare time to revise. But the time was not enough for writing an argumentative essay. I was able to extend the writing with random vocabulary in narrative. However, in writing an argumentative essay, I felt that I needed to use more sophisticated vocabulary, and thinking about vocabulary consumed a lot of time."

Excerpt 4. Advanced student in long-timed group, #47

"In writing the argumentative essay, I was not able to take enough time to brainstorm ideas. It was okay to write 300 words in one hour, but I wanted to have 20 more minutes to write out my argument, supporting ideas, and examples. On the other hand, in writing the narrative, the given time was enough because the topic was about myself. So I did not take a lot of time for brainstorming. And the word choice in narrative writing is more free than that in argumentative essays, so I can write faster in a narrative."

Table 26

Descriptive Statistics: Writing Difficulty Ratings in the Four Conditions

	High-intermediate short-timed (N = 30)		High-intermediate long-timed (N = 30)		Advanced short-timed (N = 33)		Advanced long-timed (N = 30)	
	M (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI
Q1	4.53 (2.01)	3.77, 5.28	4.33 (1.98)	3.59, 5.08	3.90 (1.75)	3.29, 4.53	4.00 (1.46)	3.45, 4.54
Q2	5.83 (1.46)	5.28, 6.38	5.50 (1.81)	4.82, 6.18	4.84 (2.00)	4.14, 5.56	4.27 (1.86)	3.57, 4.95
Q3	4.88 (1.81)	4.20, 5.55	4.46 (1.70)	3.83, 5.10	4.27 (1.55)	3.72, 4.82	4.83 (1.53)	4.26, 5.40
Q4	5.90 (1.67)	5.28, 6.52	5.03 (1.56)	4.45, 5.61	5.00 (1.92)	4.32, 5.68	4.47 (1.87)	3.77, 5.16
Q5	3.58 (1.79)	2.92, 4.25	4.07 (1.95)	3.34, 4.79	3.51 (2.15)	2.75, 4.28	4.13 (1.93)	3.41, 4.85
Q6	4.07 (1.85)	3.38, 4.76	4.27 (1.57)	3.67, 4.85	4.24 (2.06)	3.51, 4.97	4.66 (2.00)	3.94, 5.40
Q7	5.73 (2.20)	5.03, 6.59	4.30 (2.42)	3.40, 5.20	4.58 (2.25)	3.77, 5.37	2.80 (1.91)	2.08, 3.51
Q8	5.85 (2.45)	4.93, 6.76	4.20 (2.10)	3.41, 4.99	5.39 (2.45)	4.53, 6.26	3.36 (2.30)	2.51, 4.23

Note. Ratings are on a 9-point scale: 1 = strongly agree/not difficult/not interesting/not anxious/not at all, and 9 = strongly disagree/very difficult/very interesting/very anxious/a lot.

Table 27

Task Difficulty Ratings in the Four Task Conditions (One-Way ANOVA)

Item	*F*	*p*	*d*	Comparison
Q1. How difficult was the narrative essay to write?	.784	.505	.25	
Q2. How difficult was the argumentative essay to write?	4.521	.005*	.54	
Q3. I did well writing the narrative essay.	.988	.401	.22	
Q4. I did well writing the argumentative essay.	3.412	.020	.47	
Q5. How interesting was it to write the narrative essay?	.822	.484	.16	
Q6. How interesting was it to write the argumentative essay?	.533	.660	.17	
Q7. How anxious were you about the time pressure when writing the essays?	8.976	<.001*	.82	HS > AL
Q8. How much did the time limit (30 minutes/60 minutes) affect your writing?	7.113	<.001*	.64	HS > AL

*$p < .0062$ (Bonferroni adjustment)
Note. HS: high-intermediate short-timed, HL: high-intermediate long-timed, AS: advanced short-timed, AL: advanced long-timed

Table 27 presents the results of a one-way ANOVA on the difficulty ratings in the four conditions. Statistically significant differences across the four conditions (groups) were found for three out of eight items, related to the difficulty of writing argumentative essays (Q2), anxiety about time pressure (Q7), and the perception of time constraints (Q8).

The responses to Q2, regarding the difficulty of argumentative essays, showed a small effect size. For this question, the post hoc comparisons using the Bonferroni test did not show significant mean differences across the groups. However, compared to the advanced long-timed group students, the high-intermediate students in both short-timed and long-timed groups felt more difficulty in writing argumentative essays.

Q7 is related to how anxious the participants felt about the time pressure during the writing tasks. Statistically significant differences across the four conditions (groups) were found

with a medium effect size. A Bonferroni test showed that the high-intermediate short-timed group was more anxious about the time pressure than the advanced long-timed group. Compared to the advanced long-timed group, the high intermediate group felt significantly more anxious about the time pressure during the writing.

Q8 concerns the participants' perceptions of the effect of time constraints on their writing. It significantly distinguished the groups, with a medium effect size. Within the same proficiency groups, the short-timed group perceived a significantly larger effect of time constraints than the long-timed group. In addition, a significant mean difference was detected between the high-intermediate short-timed group and the advanced long-timed group. The students in the high-intermediate short-timed group believed that the allotted time had a greater effect on their writing than did those in the advanced long-timed group.

CHAPTER 5. DISCUSSION

5.1. Overview of research questions and results

As previously described, there is a growing interest in exploring writing fluency behaviors such as pausing and revising because of concerns regarding the validity of assessments of writing fluency. Instead of focusing solely on length-based measures (i.e., product-based measures) that do not consider how the writing is produced, this study explored L2 learners' writing fluency behaviors and the cognitive processes behind them. Contributing to and extending existing research on cognitive processes associated with pausing and revision behaviors, the study examined how different aspects of writing tasks, such as genre and time constraints, affect different proficiency L2 learners' writing fluency behaviors and linguistic outcomes in hopes of better understanding L2 writing. To address the first research question, the study compares overall writing fluency behaviors and linguistic outcomes in two different genres when L2 learners of different English proficiency levels write under two different time constraints. The second research question is addressed by an analysis of the participants' stimulated recall comments regarding the effects of the time constraints and the genres on their writing processes. For the third research question, the study examines how proficiency and time constraints affect writing quality (writing scores) in the two genres. The fourth research question guides the study's exploration of which writing fluency measures are related to text quality and linguistic complexity, and to what extent. The fifth research question inquires into how the L2 learners perceived the effects of the time constraints and genres after finishing the two writing

tasks. Table 28 summarizes the findings of the study. In this chapter, the five research questions

and then the overall contribution of the results of the dissertation research are discussed.

Table 28
Summary of Findings

| | | Dependent Variables | | |
Independent variables		Writing fluency behaviors	Linguistic complexity	Writing quality
Time constraint	Short-timed	The short-timed writing showed higher writing fluency		
	Long-timed		No difference	No difference
Genre	Narrative	Narrative writing showed higher fluency	Argumentative writing showed higher complexity	
	Argumentative			N/A
Proficiency	Advanced	Advanced learners showed higher writing fluency	Advanced learners showed higher writing complexity	Advanced learners showed higher writing quality
	Intermediate			

5.2. Research question 1: To what extent do proficiency and time constraints affect writing fluency behaviors and linguistic outcomes of L2 writers' writing in two genres?

The effect of time constraints in the current study appeared only in writing fluency

behavior measures. Specifically, the short-timed groups showed more fluency and less pausing

than the long-timed groups. These findings are different from those of Elder et al. (2009), who

did not find an effect from their two time constraint conditions (30 minutes and 55 minutes).

However, the previous study used fluency ratings by two raters instead of fluency measures,

which might explain the different findings. The results of the current study also differ from those

of previous studies that did find an effect of time constraints on fluency (Knoch & Elder, 2010;

Wu & Erlam 2016). Wu and Erlam (2016) compared a long-timed condition and a short-timed

condition (70% of the time the learners used in the untimed condition) and reported that the learners produced more words in their short-timed essays. Knoch and Elder (2010) also measured fluency as the number of words and found that a long-timed group (55 minutes) showed better performance than a short-timed group (30 minutes). The discrepancy between these two studies' findings and the current study's results may be due to the difference in the operationalization of fluency.

There was no effect of time constraints on linguistic outcomes. Previous research has reported mixed results for an effect of time constraints on linguistic complexity (Knoch & Elder, 2010; Wu & Erlam; 2016). Similar to the current study, Wu and Erlam (2016) found no difference between essays produced under two time constraints in terms of complexity and accuracy. However, Knoch and Elder (2010), who compared essays written in two time constraint conditions in terms of both grammatical complexity and lexical complexity, found that the short-timed condition (30 minutes) led to higher grammatical complexity than the long-timed condition (55 minutes) but did not find a difference in lexical complexity. Because they did not report their participants' L2 proficiency, however, it is difficult to compare their findings and those of the current study. In addition, they used only one grammatical complexity measure (clauses per t-unit), which may not show a full picture of linguistic complexity, considering that syntactic complexity is a multidimensional construct (Norris & Ortega, 2009).

The L2 learners' writing fluency behaviors and the linguistic complexity of their essays differed depending on the genre in which they were writing. These findings corroborate previous studies' findings of genre effects on L2 learners' writing, and more specifically on linguistic complexity (Biber & Conrad, 2009; Lu, 2011) and fluency (Beauvais et al., 2011; Medimorec & Risko, 2017; Qin & Uccelli, 2016; Van Hell et al., 2008). In the current study, a genre effect was

found in one length-based measure (MLC), one particular structure measure (CN/T), two lexical sophistication measures (WL and WF), and three fluency measures (product: words per minute, P-burst length, and number of R-bursts). These measures indicated that the learners in this study showed greater linguistic complexity but less fluency in argumentative essays than in narrative essays. These findings are similar to those of some previous studies that also found greater linguistic complexity but less fluency in argumentative essays than in narratives (Beers & Nagy, 2009; Qin & Uccelli, 2016). One explanation for this pattern is that dealing with the more demanding task (i.e., argumentative essay writing) inhibits revision behavior due to the limited availability of cognitive resources (Leijten et al., 2010; Schilperoord, 2002; Van Waes et al., 2010). Taken together, these studies' results might reflect that learners pause more when writing argumentative essays in order to engage in deeper lexical selection (i.e., searching for more sophisticated and less frequent vocabulary) as well as more complex ideas and produce more complex syntactic structures. In other words, producing language appropriate to the argumentative genre may require greater cognitive effort than producing language appropriate to the narrative genre. Hence, learners may slow their production down as they utilize more time for planning or translation, meeting the genre requirements at the expense of fluency (Beauvais et al., 2011; Kellogg, 2001).

However, the findings differ partially from those of previous studies that found a genre effect on complexity but not fluency (e.g., Yoon & Polio, 2017). The difference may be due to the measurements of fluency or the L2 proficiency of the participants. Although the current study used different writing fluency behavior measures to assess fluency and reported the L2 learners' standardized test scores and cloze-test scores, the previous studies used a traditional measure (i.e., the number of words produced in a given time) and did not use standardized test scores for

97

measuring L2 English proficiency. A contrasting result was reported by Yang (2014), who compared four genres (narrative, expository, expo-argumentative, and argumentative) with regard to complexity, accuracy, and fluency; she found higher complexity and fluency in argumentative essays than in narratives. However, she operationalized fluency as the total number of words per essay, which is a traditional length measure, whereas the present study included writing fluency behavior measures. In addition, Yang used the same cloze-test that the current study used, but her participants' mean scores on the cloze test (argumentative group: $M = 26.65$; narrative group: $M = 28.02$) were lower than those in the current study.

An interaction between time constraint and genre was found in terms of pausing between words. The L2 learners' pausing patterns differed in the two genres and in the two time-constraint conditions. In the short-timed condition, the learners paused more between words when writing argumentative essays than when writing narrative essays, whereas in the long-timed condition, they paused between words more often when writing narratives. Drawing on Kellogg's (1996) model, it was predicted that the L2 learners' fluency behaviors would differ because the amount of allowed time for a task and the requirements of a task can influence how long L2 learners stay at the translation stage and how they allocate processing time and cognitive effort for planning, translating, and monitoring. In a short-timed condition, increased time pressure may prevent smooth and responsive writing behavior, particularly for argumentative writing; however, in a long-timed condition, L2 learners may pause more while producing narratives to search for elaborate lexical items, to plan the narrative's storyline or to review their narratives as they extend the discourse with the help of extra time.

Previous research found proficiency effects for both linguistic complexity (Lu, 2011; Ortega, 2003; Wolfe-Quintero et al., 1998) and fluency (Sasaki, 2004; Van Waes & Leijten,

2015; Way et al., 2000; Yang, 2014). In the current research, a proficiency effect was found in linguistic complexity and in two of the fluency measures (product: words per minute and process: words per minute). Advanced learners produced more words per minute, reflecting their more highly developed language skills. The quantitative results did not show a proficiency effect in revision behaviors, however (e.g., number of R-bursts). This result is dissimilar to Barkaoui's (2016) finding that low proficiency learners revised more often than high proficiency learners. This difference may be due to participant factors. In the current study, the participants were all post-secondary students at the same university, who differed only in their English proficiency; in contrast, in Barkaoui's study the participants were first- or second-year graduate or undergraduate students (the high group) and pre-admission students enrolled in pre-academic ESL courses (the low group). The different writing experiences of these two groups of participants may have led to their use of different revision strategies.

No interaction between genre and proficiency was found in this study, although genre and proficiency individually affected writing fluency behaviors and linguistic outcomes. This differs from Jeong's (2017) study, which found a genre bias in proficiency. Jeong reported that novice learners performed better in the narrative genre than the expository genre, while advanced proficiency learners demonstrated better performance in the expository genre than in the narrative genre, in terms of essay scores. However, the findings of the current study indicate that both high-intermediate and advanced proficiency groups appeared to have genre awareness and understand the need to write differently in different genres (e.g., Biber & Conrad, 2009; Biber et al., 2011). In addition, because comparing narrative and argumentative essay scores is akin to comparing apples and oranges, the present study instead compared fluency behaviors and

linguistic outcomes in the two genres, which were assessed by two different rubrics; thus, the findings did not show an interaction between genre and proficiency in terms of writing quality.

5.3. Research Question 2: As evidenced by the stimulated recall data, to what extent do proficiency and time constraints affect L2 writers' writing process in the two genres?

The stimulated recall data demonstrate that there were differences in the processes underlying pausing behaviors in the two time-constraint conditions. The learners in the short-timed group spent more time in planning, which was a driving force in enhancing fluency (Sasaki, 2000). In addition, even though both time constraint groups focused more on formulation (planning and translation) than monitoring, the long-timed groups tended to spend more time on monitoring than the short-timed groups. This behavior may have resulted in the long-timed groups' lower fluency. This finding supports Kellogg's (1996) model, in the sense that it suggests that the time pressure on the short-timed group limited central executive functions, leading the learners to prioritize formulation over monitoring.

The stimulated recall data also show that the genres caused some differences in pausing behaviors; this finding is partly consistent with Kellogg's model (1996). Although the distribution of stimulated recall comments about translation processes is similar in the two time conditions, the distribution of comments about planning and monitoring during pauses differs in the two genres. Overall, the L2 learners spent more time on planning and monitoring in argumentative essays than narratives. This study's stimulated recall data show that a higher percentage of pausing comments referred to planning than to translation and monitoring across all groups when they were writing in the argumentative genre. This finding is in line with

previous research claims that the argumentative genre is more cognitively demanding and requires more planning than the narrative genre (Beauvais et al., 2011; Kellogg, 2001; Van Hell et al., 2008).

The stimulated recall data further show that more time was spent pausing between words by the short-timed groups for the argumentative essays, and by the long-timed groups for the narratives. One possible explanation for these patterns is that the combination of time pressure and the greater cognitive demand of argumentative essays required more pauses (Kellogg, 2001). In contrast, for the long-timed group, a lack of time pressure when writing narratives might have tempted the learners to do more brainstorming to extend their writing. As for the kinds of processing the learners were doing during the pauses between words, the stimulated recall data indicate that the percentages of the various writing processes (planning, translation, and monitoring) were similar between the two genres across time constraint conditions.

With regard to how their comments about revision aligned with textual locations, the learners also showed similar patterns in both genres. With regard to the overall writing processes underlying pausing and revision behaviors, Kellogg's (1996) model suggests that writing requires lower and higher cognitive processes. Many of the participants explained the pauses they made between words and between clauses, and sometimes between sentences, with comments associated with planning. These findings are similar to those of previous research in suggesting that pausing at higher text units, such as sentences, is more likely to be related to higher-level writing processes, such as planning (Révész, Kourtali, & Mazgutova, 2017; Schilperoord, 1996). Most of the learners' comments related to translation were at the word and below the word level. With regard to how their comments about revision aligned with textual locations, the learners also showed similar patterns in both genres. Most of the learners'

comments related to translation were at the word and below the word level. These patterns may be similar to their pausing behaviors in that they suggest the writers focused on retrieving lexical items or syntactic structures at this smaller discourse unit level.

Similar to previous research (e.g., Stevenson et al., 2006), the current study's stimulated recall data on writing behaviors, such as pausing and revision, also found proficiency differences. In writing the narratives, the advanced learners made more translation-related comments at the word and below the clause level than did the high-intermediate learners. In the argumentative essays, the advanced learners again made more translation-related comments at the word level than did the high-intermediate learners. Compared to the high-intermediate learners, the advanced learners also showed more translation-related revision behaviors at the word and below the clause levels. Considering that the quantitative results showed that the advanced learners produced more syntactic complexity with greater fluency than the high-intermediate learners, the advanced learners may have focused on refining syntactic structures during revision processes at the word or clause level while writing (Stevenson et al., 2006).

As for the writing processes underlying the participants' revision behaviors, only proficiency had a notable effect on them; time constraints and genres did not seem to affect the writing processes underlying the revision behaviors. These results are possibly due to the specific aspect of revision in question, in that the learners tended to focus mainly on a refining process that may be affected more by proficiency than by other factors. According to their comments, the advanced learners tended to spend more time on translation processes (about 60–70%) than did the high-intermediate learners (about 40–50%), as a larger number of pausing and revision comments about lexical retrieval, syntactic encoding, and cohesion were made by the advanced learners than by the high-intermediate learners. Possibly, the amount of engagement in

translation processes underlying revision behaviors might have contributed to the differences in production at the two proficiency levels. This finding is dissimilar to the findings of previous L2 research that utilized keystroke logging (e.g., Barkaoui, 2016; Stevenson et al., 2006). Barkaoui (2016) found that low proficiency learners made significantly more revisions than high proficiency learners, and Stevenson et al. (2006) did not find differences between their two proficiency groups; however, they divided the two groups by relative proficiency instead of using standardized scores. These findings in the current study are similar to those of Sasaki (2000), who observed that expert writers spend more time on rhetorical refining than novices do. In the present study, the advanced learners devoted more time to translation processes including retrieving words, syntactic encoding, and cohesion than did the high-intermediate learners.

5.4. Research Question 3: How do L2 proficiency and time constraints affect writing quality in two essay genres?

Only proficiency had a significant effect on writing quality; this was true for both the narrative and argumentative essays. Previous studies have also found proficiency effects on quality (e.g., Jeong, 2017; Xu & Ding, 2014). Possibly, more advanced learners' greater ability to manage the various necessary writing processes allows them to produce higher-quality writing according to all five scales, that is, content, organization, vocabulary, language use, and mechanics (Chenoweth & Hayes, 2001). In the current study, the stimulated recall data show that, compared to the high-intermediate learners, the advanced learners engaged relatively more in translation processes than in planning processes. This finding suggests that greater English proficiency may enable learners to pay more attention to form during writing. In addition, for the

high-intermediate learners, who can be assumed to have had less L2 experience than the advanced learners, retrieving lexical items likely required more effort. Thus, in addition to the fact that advanced learners know more language than high-intermediate learners, the extent to which learners engage in different writing processes may affect writing quality (Kellogg, 1990).

In contrast to L2 proficiency, the time constraints did not affect writing quality. The findings of this study are similar to those of other previous studies that did not find significant effects of time constraints on writing quality (Caudery, 1990; Elder et al., 2009; Knoch & Elder, 2010; Powers & Fowles, 1996). Based on the current study's stimulated recall data, the short-timed and long-timed groups did not employ noticeably different writing processes in the two different time conditions, except for in planning and monitoring processes, and these differences were not reflected in writing quality. However, these results are different from those of prior studies that have found longer-timed groups to produce higher quality writing (Hale, 1992; Wu & Erlam, 2016). Hale (1992) suggested that the addition of 15 minutes increased mean scores by one-third of the standard deviation; however, it is unclear whether Hale's results demonstrate that adding 15 minutes actually contributed to increases in mean scores. Wu and Erlam (2016) compared rated scores on task achievement, coherence and cohesion, lexical variation, grammatical range and accuracy, and overall quality between two time conditions. They found a slightly significant difference ($p = .04$) only in task achievement (content), which implies that time constraints did not affect writing quality much in their study.

5.5. Research Question 4: Which fluency measures are related to text quality and linguistic complexity, and to what extent?

Writing fluency measures were found to be related to writing quality in both genres. These results are similar to previous research results that have shown a relationship between writing fluency and quality (e.g., Barkaoui & Knouzi, 2018; Beauvais et al., 2011; Spelman et al., 2008; Stevenson et al., 2006). There are, however, some key differences. Namely, unlike the current study, Barkaoui and Knouzi (2018) operationalized fluency as the number of words, and Spelman et al. (2008) found a relationship between text length and quality. In the current study, writing fluency measures including pausing behaviors underlying different writing processes were positively associated with writing quality. However, no relationship between revision behaviors and writing quality was found. One possible implication of these results is that the extent to which learners have automatized their writing processes, such as how rapidly they can retrieve vocabulary from long-term memory, may affect writing quality (Kellogg, 1990). Nevertheless, based on the relationship between writing fluency and quality, it suffices to say that writing fluency could be good indicators of L2 learners' writing quality.

Writing quality was best predicted by different fluency measures depending on genre. The findings are similar to those of Qin and Uccelli (2016), who used length and lexical, syntactic, and discourse features to see which measures predicted writing quality in narrative and argumentative essays, and found length to be the most predictive of quality. In this study, in the narrative essays, writing quality was best predicted by an increase in product: words per minute and R-burst length; however, in the argumentative essays, writing quality was best predicted by an increase in process: words per minute and a decrease in one revision measure (i.e., the number

of R-bursts). Therefore, both this study and Qin and Uccelli's indicate that fluency measures predict writing quality differently in the two genres.

As was expected based on previous research that showed the relationship between complexity and fluency (e.g., Foster & Skehan, 1996; Oh, 2006), the fluency measures were related to the complexity measures in the present study. However, the relationship between fluency measures and complexity measures differed depending on genre. Again, this is similar to Qin and Uccelli's findings (2016). The results confirmed that the complexity and fluency constructs can measure different dimensions of L2 performance in different writing tasks (Housen & Kuiken, 2009; Housen et al., 2012). In addition, the correlations found in this study confirm the assumptions that writing fluency behaviors are related to linguistic complexity, and indicate that these constructs may share an underlying dimension (Medimorec & Risko, 2017).

5.6. Research Question 5: How do L2 writers perceive the effects of time constraints and genre on their writing?

As for the learners' perceptions of the effect of the time constraints, more than half of the learners in the long-timed groups considered the time enough for both genres, which echoes the findings of the previous research (e.g., Knoch & Elder, 2010; Powers & Fowles, 1996). And while they also largely perceived the short-timed conditions as insufficient, differences between the two time-constraint groups were detected only in writing fluency. No difference arising from the time constraints was found in linguistic complexity or writing quality. In other words, the shorter time seemed to elicit more fluent language without negatively affecting linguistic outcomes and writing quality.

The students in the high-intermediate short-timed group believed that the allotted time affected their writing more than those in the advanced long-timed group. For high-intermediate learners, time pressure may increase anxiety (Weigle, 2002), which in turn could affect their writing performance. However, as the comparison between high-intermediate short-timed and long-timed groups showed, differences in linguistic complexity and quality arising from the time constraints were minimal.

With regard to the perceived difficulty of writing essays in the two genres, more than half of the learners in both proficiency groups perceived the argumentative essay to be more difficult to write than the narrative, as previous studies have suggested (Ruiz-Funes, 2014, 2015). This is presumably due to the greater cognitive demands of the argumentative essay and the different functional demands of the two genres (Biber et al., 2011; Leijten et al., 2010; Van Waes et al., 2010). For this study, it is difficult to tease apart possible effects of the functional demands versus the cognitive demands of genres on the learners' perceptions of difficulty (e.g., Yoon & Polio, 2017). The survey results suggested that the learners in both proficiency groups tried to use more sophisticated vocabulary and structures in argumentative essays than in narratives. The findings of higher linguistic complexity in the argumentative essays than in the narratives and of lower fluency in the argumentative essays than in the narrative essays align with these survey results.

5.7. Contributions of this dissertation

5.7.1. Understanding time constraints

In this study, time constraints had minimal effects on L2 writing products in terms of linguistic complexity and writing quality. Providing L2 learners with extra time to plan and edit their language was expected to make a difference, but the additional 30 minutes of the long-timed condition did not contribute to increased complexity or quality. Although this study could not provide the learners with unlimited time for logistical reasons, the long-timed condition doubled the short-timed condition in order to mimic the untimed conditions in which academic writing is typically done. Among the previous studies (e.g., Knoch & Elder, 2010) that did not find significant differences between two time-constraint conditions as the current study does, Caudery (1990) provided a range of possible explanations for his null findings; some of these explanations may help understand the findings of the current study. Among the possible explanations he suggested, the participants' level of training in writing skills may have contributed to the finding of no difference in the linguistic complexity and quality of their writing. Because all of the L2 learners in the current study had received sufficient training to have gotten high standardized test scores for writing, they clearly had had much practice in writing timed essays; in short, it is possible that the 30 extra minutes did not lead to differences in linguistic complexity and writing quality because the learners had practiced short-timed writing and knew how to use various strategies for writing under time constraints.

Although the two distinct time-constraint groups showed no differences in linguistic complexity and quality, effects of time constraints on the writing process were seen in the fluency measures and stimulated recall. Different time constraints led to differences in how the

L2 learners planned and edited their writing, which were reflected in the writing fluency behavior measures. The L2 learners in the short-timed group showed more fluent writing behavior than those in the long-timed group. The stimulated-recall data revealed differences in the processes underlying pausing behaviors in the two time-constraint conditions. The learners in the short-timed conditions more often used their pauses to plan their writing than the learners in the long-timed conditions did; this additional planning may have allowed the learners in the short-timed condition to use less cognitive effort for transcription, which in turn may have led to the higher fluency in the short-timed condition than in the long-timed condition. In other words, the difference in writing processes seemed to contribute to the difference in writing fluency behaviors.

5.7.2. Understanding fluency: Genre effects and fluency's relationship with complexity and writing quality

The genre effect was evident in linguistic complexity; genre effects showed up in both writing processes and writing products as detected through fluency measures and stimulated recall. In addition to the effects of genre on linguistic complexity, which previous research also has found, the L2 learners' writing fluency behaviors differed in the two genres in the current study. The differences arising from complexity and fluency were confirmed by the stimulated recall data. When the learners wrote argumentative essays, their planning-related pauses were more frequent than when they wrote narratives. During such pauses, the learners planned the content and organization of their writing, which are especially important in argumentative

essays. Thus, specific pausing behaviors and the processes in which learners engage during the pauses may contribute to differences in complexity and fluency in the two genres.

The current study found that writing fluency was related to writing quality in both genres. Specifically, process: words per minute, product: words per minute, and p-burst length were significantly related to writing quality in the two genres. Many other empirical studies (e.g., Beers & Nagy, 2009; Qin & Uccelli, 2016) have also investigated the relationship between linguistic features and writing quality in different genres, but they considered writing fluency as length of writing. However, considering fluency to be a multidimensional construct, the current study tried to better elucidate the relationship of writing fluency and quality by adding process-based measures instead of looking at only one length-based measure. Based on the results of the different fluency measures, the study suggests that writing fluency features can be indicative of writing quality.

A relationship between fluency measures and complexity measures was also found. As one of the CAF measures, fluency is believed to relate to complexity and accuracy. The current study provides empirical evidence for the relationship between the two constructs of complexity and fluency in both of the two genres. Oh (2006) empirically tested the relationship between the two constructs, but she operationalized fluency as the number of T-units and the number of clauses and examined only argumentative essays in testing settings. This study, unlike such previous research, looked at the fluency construct multidimensionally by employing both process-based and product-based measures to examine the relationship between the fluency and complexity constructs in the two genres.

L2 learners' perceptions of genres can also play a role in their writing processes and products. More than half of the learners in both proficiency groups in this study perceived the

argumentative essay to be more difficult to write than the narrative essay. The reasons behind these differences in perceived difficulty are mainly due to the structure and the language of argumentative writing. The learners had learned genre differences from English academic writing classes, but they still struggled with the specific requirements of argumentative writing, such as the need to provide clear arguments, supporting ideas, and appropriate examples. In addition, the survey suggested that the learners tried to use more sophisticated vocabulary and structures in argumentative essays than in narrative essays. Their different perceptions of the difficulty of the two genres seemed to be reflected in their writing processes and products: higher linguistic complexity and lower writing fluency behaviors in argumentative essays than in narratives.

CHAPTER 6. CONCLUSION

6.1. Summary

This dissertation sought to advance the L2 writing research on writing fluency behaviors, and sheds light on the interplay of time constraints, genre, and proficiency in L2 writing fluency behaviors and linguistic outcomes. The study reached several conclusions. First, L2 learners produced more complex language and showed less fluent writing behaviors in argumentative than in narrative essays. A significant interaction between genre and time condition was found in the number of pauses between words. A proficiency effect was found in linguistic complexity and fluency, while a time constraint effect was detected only in fluency and writing fluency behaviors. Second, the stimulated recall data indicate that the learners tended to spend more time on planning and monitoring in the argumentative genre than the narrative genre. For the time constraint comparisons, the short-timed groups did more planning than the long-timed groups, whereas the long-timed groups tended to spend more time on monitoring than the short-timed groups. The advanced learners' comments about pausing and revision are more associated with translation than the intermediate learners' for both narrative and argumentative essays. As for writing processes according to locations, higher textual units are associated with higher level processes. Third, L2 proficiency affected writing quality in both genres; however, the difference in time conditions (30 minutes vs. 60 minutes) did not affect writing quality. Fourth, writing fluency measures were correlated with linguistic complexity and writing quality; however, these correlations differed by genre. Writing fluency and revision behaviors can predict writing quality in both genres. Fifth, more than half of the learners in each group perceived argumentative essays as more difficult to write than narratives. The learners perceived the time constraints

differently; more than half of the learners in the long-timed group considered the time allotment enough for both genres. Compared to learners in the advanced long-timed group, learners in the high-intermediate short-timed group felt more anxiety and believed more strongly that writing time affected their writing quality.

6.2. Theoretical, methodological, and pedagogical implications

From a theoretical perspective, the results shed light on how different genres and time constraints affect different proficiency learners' writing fluency behaviors and linguistic outcomes, in terms of process and production. Based on Kellogg's (1996) writing model, this study provided empirical evidence as to how L2 learners of different proficiencies may show different cognitive processes underlying writing behaviors, writing fluency behaviors, language complexity, and text quality when writing in different genres under varying time constraints. This study was intended to help explain cognitive processes associated with writing behaviors in different writing genres and time constraints. In addition, previous research has tended to focus on production differences associated with genres, time constraints, and proficiency. However, this study delved into what leads to these differences by investigating the writing fluency behaviors underlying writing processes in addition to investigating production.

With regard to methodological implications, this dissertation research included keystroke logging to unobtrusively capture L2 writing behaviors such as pausing and revising. Along with keystroke logging and automatic textual analysis, the study also employed stimulated recall protocols to enable cognitive-linguistic analysis of writing processes. As it examined writing processes and products multidimensionally, the study used a combination of research methods in

order to achieve more valid and accurate interpretations of different proficiency L2 learners' writing processes when they responded to different genre prompts under different time constraints.

This study holds pedagogical and assessment implications. With respect to L2 writing instruction, teachers tend to present different genres to students and set different time constraints for assignments. In this study, the differences that arise from genres and time constraints were explained in terms of L2 learners' writing processes and production. For instance, in writing argumentative essays, the L2 learners in this study made more planning-related comments than translation- and monitoring-related comments when they were explaining why they paused. These findings indicate that learners may benefit from learning about different planning strategies for writing argumentative essays. In particular, the findings of this study are crucial for test developers and teachers for designing writing tests and assignments. The findings showed that the effects of time constraints (30 minutes vs. 60 minutes) on the written product, in term of quality and language, were not significant; however, the students felt more anxiety in writing short-timed essays than long-timed essays. Moreover, keystroke logging and replaying keystroke logging can provide teachers with insights when diagnosing students' difficulties in writing. For instance, information obtained from keystroke logging and surveys together can help teachers understand which time constraints or tasks are appropriate for their students at different levels. In addition, from the learners' perspective, as Ranalli et al. (2018) demonstrated, keystroke logging can also give students information about their writing process. As they read their own keystroke logging information, learners can become more aware of the cognitive processes underlying their writing fluency behaviors.

6.3. Limitations and future research

The limitations of this study should be acknowledged. As described in the method section, the advanced learners showed better keyboarding skills than the high-intermediate learners. It is clear that proficiency affected their writing processes and products based on the results of the study such as those regarding writing quality; however, keyboarding skill differences in the two English proficiency groups might also have contributed to writing fluency behavior differences (e.g., Barkaoui & Knouzi, 2018).

In addition, this study manipulated time constraints as 30-minute and 60-minute conditions. For logistical reasons, the longer-timed condition was used to mimic an untimed condition; however, this manipulation may lack authenticity. Based on the survey, some of the learners in the long-timed groups still felt the allotted time was not enough for writing in either genre.

Following previous research (e.g., Révész, Kourtali, & Mazgutova, 2017; Spelman Miller et al., 2008), this study used a threshold of two seconds for determining pauses. However, some researchers have suggested that different thresholds for pauses such as 200 milliseconds or 500 milliseconds might capture different dimensions of writing fluency behaviors such as lower levels of writing processes (Van Waes & Leijten, 2015; Wengelin, 2006).

Future research should examine changes in writing fluency behaviors and linguistic outcomes longitudinally (e.g., Spelman Miller et al., 2008). This study used a within-groups design for genre and a between-groups design for exploring proficiency and time constraint effects. The study found time constraint, genre, and proficiency effects on writing fluency behaviors and linguistic outcomes. However, if a study investigated how learners write in the

two genres under different time constraints over time, the findings might be different from those of this study, because learners can be expected to develop their L2 over time.

In addition, although the current study showed the relationship between complexity and fluency, it is open to question whether accuracy, one of the CAF measures, is related to complexity and fluency. The current study did not include accuracy measures in the analysis because the measure may not be particularly useful for assessing L2 development or differentiating learners by proficiency (e.g., Lambert & Kormos, 2014). However, for the purpose of theory building, it may be useful to add accuracy measures to shed light on the relationship between accuracy and fluency in writing.

Some recent studies have employed eye-tracking technology as well as stimulated recall and key-stroke logging (e.g., Ranalli et al., 2018; Révész et al., in press). Eye-tracking methods might uncover other cognitive processes underlying writing behaviors. However, the low frequency eye-trackers such as Tobii 60x, which usually do not hamper natural writing behaviors, are less accurate than high frequency eye-trackers such as Eyelink 1000, and the data from the eye-trackers are messy. Although eye-trackers such as Eyelink 1000 are very accurate in assessing learners' saccades during writing, it is almost impossible to get participants to act naturally because they need to keep their heads still on a chin-rest to assure high tracking accuracy. Nevertheless, when highly accurate eye-tracking technology that does not intervene in the natural writing process becomes available, it will be helpful for future investigations of learners' writing processes.

APPENDICES

APPENDIX A: Prompts for the Narrative and the Argumentative Essays (Yoon, 2017)

Narrative prompt: Your friend has plans to learn a foreign language but is afraid it might be useless to spend the time learning a language. You have successfully learned a foreign language and use it often. You want to show your friend that language learning and use can be interesting by telling him/her about your positive experience. Tell a story about one of your positive experiences related to foreign language use. Be sure to fully develop your story by including specific details.

Argumentative prompt: You attended a seminar and the main theme was that using a foreign language fluently has become necessary in this globalized era. Write an essay about whether you agree or disagree with the statement about the necessity of foreign language abilities. Support your position with reasons. Be sure to fully develop your essay by including clear explanations and logical supporting ideas.

DIRECTIONS

1. Read the passage quickly to get the general meaning.

2. Write **only one word** in each blank next to the item number. Contractions are considered to be one word.

3. Check your answers.

You have 25 minutes to complete the cloze test.

EXAMPLE: The boy walked up the street. He stepped on a piece of ice. He fell (1) *down* but he didn't hurt himself.

MAN AND HIS PROGRESS

Man is the only living creature that can make and use tools. He is the most teachable of living beings, earning the name of Homo sapiens. (1)_____ ever restless brain has used the (2)_____ and the wisdom of his ancestors (3)_____ improve his way of life. Since (4)_____ is able to walk and run (5)_____ his feet, his hands have always (6)_____ free to carry and to use (7)_____ . Man's hands have served him well (8)_____ his life on earth. His development, (9)_____ can be divided into three major (10)_____ , is marked by several different ways (11)_____ life.

Up to 10,000 years ago, (12)_____ human beings lived by hunting and (13)____ . They also

picked berries and fruits, (14)_____ dug for various edible roots. Most (15)___ , the men were the

hunters, and (16)_____ women acted as food gatherers. Since (17)_____ women were

busy with the children, (18)_____ men handled the tools. In a (19)_____ hand, a dead

branch became a (20)_____ to knock down fruit or (21)_____ for tasty roots.

Sometimes, an animal (22)_____ served as a club, and a (23)_____ piece of stone, fitting

comfortably into (24)_____ hand, could be used to break (25)_____ or to throw at an animal.

(26)_____ stone was chipped against another until (27)_____ had a sharp edge. The

primitive (28)_____ who first thought of putting a (29)_____ stone at the end of a (30)

made a brilliant discovery: he (31)_____ joined two things to make a (32)_____ useful tool,

the spear. Flint, found (33)_____ many rocks, became a common cutting (34)_____ in the

Paleolithic period of man's (35)_____ . Since no wood or bone tools (36)_____ survived, we

know of this man (37)_____ his stone implements, with which he (38)_____ kill animals, cut

up the meat, (39)_____ scrape the skins, as well as (40)_____ pictures on the walls of the

(41)_____ where he lived during the winter.

(42)_____ the warmer seasons, man wandered on (43)_____ steppes of Europe without a fixed

(44)_____ , always foraging for food. Perhaps the (45)_____ carried nuts and berries in shells

(46)_____ skins or even in light, woven (47)_____ . Wherever they camped, the primitive

people (48)_____ fires by striking flint for sparks (49)_____ using dried seeds, moss, and

rotten (50)_____ for tinder. With fires that he kindled himself, man could keep wild animals

away and could cook those that he killed, as well as provide warmth and light for himself.

Answer keys

"Man and his progress" - answer keys

Exact answer Acceptable answer scoring would also include these possibilities

1 His man's, our, the

2 Knowledge, accomplishments, culture, cunning, examples, experience(s), hands, ideas, information, ingenuity, instinct, intelligence, mistakes, nature, power, skill(s), talent, teaching, technique, thought, will, wit, words, work

3 to

4 man, he

5 on, upon, using, with

6 been, felt, hung, remained

7 tools, adequately, carefully, conventionally, creatively, diligently, efficiently, freely, implements, objects, productively, readily, them, things, weapons

8 during, all, for, improving, in, through, throughout, with

9 which, also, basically, conveniently, easily, historically, however, often, since, that, thus

10 periods, areas, categories, divisions, eras, facets, groups, parts, phases, sections, stages, steps, topics, trends

11 of, for, in, through, towards

12 all, early, hungry, many, most, only, primitive, the, these

13 fishing, farming, foraging, gathering, killing, scavenging, scrounging, sleeping, trapping

14 and, often, ravenously, some, they

15 often, always, emphatically, important, nights, normally, of, times, trips

16 the, all, house, many, most, older, their, younger

17 the, all, many, married, most, often, older, primate, these

18 the, all, constructive, many, most, older, primate, tough, younger

19 man's, able, big, closed, coordinated, creative, deft, empty, free, human('s), hunter's,
learned, needed, needy, person's, right, single, skilled, skillful, small, strong, trained

20 tool, club, device, instrument, pole, rod, spear, stick, weapon

21 dig, burrow, excavate, probe, search, test

22 bone, arm, easily, foot, head, hide, horn, leg, skull, tail, tusk

23 sharp, big, chipped, fashioned, flat, hard, heavy, large, rough, round, shaped, sizeable, small,
smooth, soft, solid, strong, thin

24 the, a, his, man's, one('s)

25 nuts, apart, bark, bones, branches, coconuts, down, firewood, food, heads, ice, items, meat,
objects, open, rocks, shells, sticks, stone, things, tinder, trees, wood

26 one, a, each, flat, flint, glass, hard, obsidian, shale, softer, some, the, then, this

27 it, each, one, they

28 man, being, creature, human, hunter, men, owner, people, person

29 sharp, glass, hard, jagged, large, lime, pointed, sharpened, small

30 stick, bone, branch, club, log, pole, rod, shaft

31 had, accidentally, cleverly, clumsily, conveniently, creatively, dexterously, double, easily,
first, ingeniously, securely, simply, soon, suddenly, tastefully, then, tightly, would

32 very, bad, extremely, good, hunter's, incredibly, intelligent, long, modern, most, necessarily,
new, portentously, quite, tremendously, useful

33 in, all, among, amongst, by, inside, on, that, using, within

34 tool, device, edge, implement, instrument, item, material, method, object, piece, practice, stone, utensil

35 development, age, ancestry, discoveries, era, evolution, existence, exploration, history, life, time

36 have, actually, apparently, ever

37 by, and, for, from, had, made, through, used, using

38 could, did, would

39 and, carefully, help, or, skillfully, then, would

40 draw, carve, create, drawing, engrave, hang, paint, painting, place, sketch, some, the

41 cave(s), animals, place(s), room

42 in, and, during, with

43 the, across, aimless, all, barren, dry, flat, high, in, long, many, plain, stone, through, to, toward, unknown, various

44 home, appetite, camp, course, destination, destiny, diet, direction, domain, foundation, habitat, income, knowledge, location, lunch, map, meal, path, pattern, place, plan, route, supplement, supply, time, weapons

45 women, children, families, group, human, hunter, man, men, people, primitives, voyager, wanderers, woman

46 or, and, animal, animal's, covered, in, like, of, on, their, using, with

47 baskets, bags, blankets, chests, cloth(s), clothes, fabric, garments, hides, material, nets, pouches, sacks

48 made, began, built, lighted, lit, produced, started, used

49 and, also, by, occasionally, or, then, together, while

50 wood, bark, branches, dung, forage, grass, leaves, lumber, roots, skin, timber, tree(s)

APPENDIX C: Timed Key-boarding Skill Test

Write the sentence below as many times as you can for two minutes.

I voluntarily agree to participate in this writing research.

APPENDIX D: Language Experience and Proficiency Questionnaire (Marian et al., 2007)

Name		Date	
Age		Gender	

Please list all the languages you know in order of dominance:

English is my _____ language. (insert ordinal number: 1st, 2nd, and so on)

All questions below refer to your knowledge of English.

Please list the number of years and months you spent in each language environment:

	Years	Months
An English-speaking country		

Please provide the following information about your TOEFL/IELTS/TOEIC:

Test:	Date taken:	Total score:

APPENDIX E: Exit Questionnaire (Adapted from Yoon, 2017)

1. How did you feel about writing narrative and argumentative essays? Is one type of essay writing more difficult than the other (in terms of brainstorming/planning, writing, and revising)? Why? Please explain.

2. How difficult was the narrative essay to write?

(Not difficult at all) 1-2-3-4-5-6-7-8-9 (Very difficult)

3. How difficult was the argumentative essay to write?

(Not difficult at all) 1-2-3-4-5-6-7-8-9 (Very difficult)

4. I did well writing the narrative essay.

(Strongly Agree) 1-2-3-4-5-6-7-8-9 (Strongly disagree)

5. I did well writing the argumentative essay.

(Strongly Agree) 1-2-3-4-5-6-7-8-9 (Strongly disagree)

6. How interesting was it to write the narrative essay?

(Very interesting) 1-2-3-4-5-6-7-8-9 (Not interesting)

7. How interesting was it to write the argumentative essay?

(Very interesting) 1-2-3-4-5-6-7-8-9 (Not interesting)

8. How anxious were you about the time pressure when writing the essays?

(Not anxious at all) 1-2-3-4-5-6-7-8-9 (Very anxious)

9. How much did the time (30 minute/1 hour) affect your writing?

(Not at all) 1-2-3-4-5-6-7-8-9 (A lot)

10. Do you think the time allotted was enough to write essays (both genres)? Please explain.

APPENDIX F: Stimulated Recall Protocol (Barkaoui, 2015 and Gass & Mackey, 2017)

As we watch the video, I'll be asking you questions about what you were doing. At times I'll even stop the video so we can examine a word choice, a revision and so forth. As you watch your writing unfold, try to recall what you were thinking at the time; try to put your mind back into the task. Anytime you remember something, say it. Interrupt me, stop the video if you want. I am interested in finding out what you were thinking when you were writing, and it doesn't matter at all to me if those thoughts were silly or profound. Again, I would like you to tell me what you were thinking when you were completing the task, NOT what you are thinking now. I will audio-record our conversation so I don't have to divide my attention by taking notes.

Open-ended questions will be used:

- What were you thinking at this point?

- Is there anything else that comes to your mind?

- I see you stopped writing. What were you thinking then?

- I see you changed the text. Can you tell me what you were thinking then?

- Can you tell me your thoughts when you paused (or made a change)?

APPENDIX G: Argumentative Essay Rubric (Connor-Linton & Polio, 2014)

	Content		Organization		Vocabulary		Language Use	Score/2	Mechanics
20 16	Thorough and logical development of thesis Substantive and detailed No irrelevant information Interesting A substantial number of words for amount of time given	20 16	Excellent overall organization Clear thesis statement Substantive introduction and conclusion Excellent use of transition word Excellent connections between paragraphs Unity within every paragraph	20 16	Very sophisticated vocabulary Excellent choice of words with no errors Excellent range of vocabulary Idiomatic and near native-like vocabulary	20 16	No major errors in word order or complex structures No errors that interfere with comprehension Only occasional errors in morphology Frequent use of complex sentences Excellent sentence variety	20 16	Appropriate layout with indented paragraphs No spelling errors No punctuation errors
15 11	Good and logical development of thesis Fairly substantive and detailed Almost no irrelevant information Somewhat interesting An adequate number of words for the amount of time given	15 11	Good overall organization Clear thesis statement Good introduction and conclusion Good use of transition words Good connections between paragraphs Unity within most paragraphs	15 11	Somewhat sophisticated vocabulary Attempts, even if not completely successful, at sophisticated vocabulary Good choice of words with some errors that don't obscure meaning Adequate range of vocabulary but some repetition Approaching academic register	15 11	Occasional errors in awkward order or complex structures Almost no errors that interfere with comprehension Attempts, even if not completely successful, at a variety of complex structures Some errors in morphology Frequent use of complex sentences Good sentence variety	15 11	Appropriate layout with indented paragraphs No more than a few spelling errors in less frequent vocabulary No more than a few punctuation errors
10 6	Some development of thesis Not much substance or detail Some irrelevant information Somewhat uninteresting Limited number of words for the amount of time given	10 6	Some general coherent organization Minimal thesis statement or main idea Minimal introduction and conclusion Occasional use of transitions words Some disjointed connections between paragraphs Some paragraphs may lack unity	10 6	Unsophisticated vocabulary Limited word choice with some errors obscuring meaning Repetitive choice of words No resemblance to academic register	10 6	Errors in word order or complex structures Some errors that interfere with comprehension Frequent errors in morphology Minimal use of complex sentences Little sentence variety	10 6	Appropriate layout with most paragraphs indented Some spelling errors in less frequent and more frequent vocabulary Several punctuation errors
5 0	No development of thesis No substance or details Substantial amount of irrelevant information Completely uninteresting Very few words for the amount of time given	5 0	No coherent organization No thesis statement or main idea No introduction and conclusion No use of transition words Disjointed connections be-tween paragraphs Paragraphs lack unity	5 0	Very simple vocabulary Severe errors in word choice that often obscure meaning No variety in word choice No resemblance to academic register	5 0	Serious errors in word order or complex structures Frequent errors that interfere with comprehension Many error in morphology Almost no attempt at complex sentences No sentence variety	5 0	No attempt to arrange essay into paragraphs Several spelling errors even in frequent vocabulary Many punctuation errors

APPENDIX H: Narrative Rubric (Adapted from Connor-Linton & Polio, 2014)

	Content		Organization		Vocabulary		Language Use	Score/2	Mechanics
20 16	Thorough and logical development of storyline Vivid and detailed No irrelevant information Interesting A substantial number of words for amount of time given	20 16	Unity within every paragraph Excellent overall organization Clear sequence of events and topic Clear sense of beginning and end Excellent use of transition word	20 16	Very sophisticated vocabulary Excellent choice of words with no errors Excellent range of vocabulary Idiomatic and near native-like vocabulary	20 16	No major errors in word order or complex structures No errors that interfere with comprehension Only occasional errors in morphology Frequent use of complex sentences Excellent sentence variety	20 16	No spelling errors No punctuation errors
15 11	Good and logical development of storyline Fairly vivid and detailed Almost no irrelevant information Somewhat interesting An adequate number of words for the amount of time given	15 11	Unity within most paragraphs Good overall organization Good sequence of events and topic Good sense of beginning and end Good use of transition words	15 11	Somewhat sophisticated vocabulary Attempts, even if not completely successful, at sophisticated vocabulary Good choice of words with some errors that don't obscure meaning Adequate range of vocabulary but some repetition	15 11	Occasional errors in awkward order or complex structures Almost no errors that interfere with comprehension Attempts, even if not completely successful, at a variety of complex structures Some errors in morphology Frequent use of complex sentences Good sentence variety	15 11	No more than a few spelling errors in less frequent vocabulary No more than a few punctuation errors
10 6	Some development of storyline Not much vividness or detail Some irrelevant information Somewhat uninteresting Limited number of words for the amount of time given	10 6	Some paragraphs may lack unity Some general coherent organization Limited sequence of events or topic Limited sense of beginning and end Occasional use of transitions words	10 6	Unsophisticated vocabulary Limited word choice with some errors obscuring meaning Repetitive choice of words	10 6	Errors in word order or complex structures Some errors that interfere with comprehension Frequent errors in morphology Minimal use of complex sentences Little sentence variety	10 6	Some spelling errors in less frequent and more frequent vocabulary Several punctuation errors
5 0	No development of storyline No vividness or details Substantial amount of irrelevant information Completely uninteresting Very few words for the amount of time given	5 0	Paragraphs lack unity No coherent organization No sequence of events or topic No sense of beginning and end No use of transition words	5 0	Very simple vocabulary Severe errors in word choice that often obscure meaning No variety in word choice	5 0	Serious errors in word order or complex structures Frequent errors that interfere with comprehension Many error in morphology Almost no attempt at complex sentences No sentence variety	5 0	Several spelling errors even in frequent vocabulary Many punctuation errors

APPENDIX I: Reasons for Pausing and Revision: Summary of Stimulated Comments

Table I-1. Number of comments for pausing in stimulated recalls (high intermediate short timed group)

	Planning			Translation				Monitoring	No recall	Unspecified	Total
	Content	Organization	Total	Lexical retrieval	Syntactic encoding	Cohesion	Total				
Narrative (*N* =2)											
Within words	0	0	0 (0 %)	0	0	0	0 (0 %)	0 (0 %)	1 (0 %)	0	1 (0 %)
Between words	21	2	23 (22 %)	22	4	0	26 (25 %)	1 (0 %)	6 (6 %)	4	60 (58 %)
Between clauses	16	3	19 (18 %)	3	1	1	5 (5 %)	9 (8 %)	0 (0 %)	1	34 (33 %)
Between sentences	2	2	4 (3 %)	0	1	0	1 (0 %)	4 (4 %)	0 (0 %)	0	9 (9 %)
Total	39	7	46 (44 %)	25	6	1	32 (31 %)	14 (13 %)	7 (7 %)	5	104 (100 %)
Argumentative (*N* =2)											
Within words	0	0	0 (0 %)	1	0	0	1 (1 %)	0 (0 %)	0 (0 %)	0	1 (1 %)
Between words	15	2	17 (19 %)	15	4	2	21 (24 %)	1 (1 %)	1 (1 %)	8	48 (54 %)
Between clauses	14	4	18 (20 %)	2	1	0	3 (3 %)	2 (2 %)	4 (4 %)	0	27 (30 %)
Between sentences	9	3	12 (13 %)	0	0	0	0 (0 %)	1 (1 %)	0 (0 %)	0	13 (15 %)
Total	38	9	47 (53 %)	18	5	2	25 (28 %)	4 (4 %)	5 (6 %)	8	89 (100 %)

Table I-2. Number of comments for revision in stimulated recalls (high intermediate short timed group)

	Planning			Translation				No recall	Unspecified	Total
	Content	Organization	Total	Lexical retrieval	Syntactic encoding	Cohesion	Total			
Narrative (*N* =2)										
Below the word	0	0	0 (0 %)	1	0	0	1 (2 %)	1 (2 %)	0	2 (3 %)
At the word level	1	0	1 (2 %)	8	10	2	20 (34 %)	0 (0 %)	3	24 (40 %)
Below the clause level	5	0	5 (8 %)	5	2	2	9 (15 %)	1 (2 %)	6	21 (36 %)
At the clause level or above	7	2	9 (15 %)	0	0	1	1 (2 %)	0 (0 %)	0	10 (17 %)
At the sentence level or above	0	2	2 (3 %)	0	0	0	0 (0 %)	0 (0 %)	0	2 (3 %)
Total	13	4	17 (29 %)	14	12	5	31 (53 %)	2 (3 %)	9	59 (100 %)
Argumentative (*N* =2)										
Below the word	1	0	1 (2 %)	2	0	0	2 (4 %)	0 (0 %)	0	3 (6 %)
At the word level	2	0	2 (4 %)	10	4	1	15 (29 %)	1 (2 %)	3	21 (40 %)
Below the clause level	5	0	5 (10 %)	4	1	0	5 (10 %)	1 (2 %)	2	13 (25 %)
At the clause level or above	5	0	5 (10 %)	2	1	0	3 (6 %)	1 (2 %)	1	10 (19 %)
At the sentence level or above	0	3	3 (6 %)	0	0	0	0 (0 %)	0 (0 %)	2	5 (10 %)
Total	13	3	16 (31 %)	18	6	1	25 (48 %)	3 (6 %)	8	52 (100 %)

Table I-3. Number of comments for pausing in stimulated recalls (high intermediate long timed group)

	Planning			Translation				Monitoring	No recall	Unspecified	Total
	Content	Organization	Total	Lexical retrieval	Syntactic encoding	Cohesion	Total				
Narrative (N =2)											
Within words	0	0	0 (0 %)	2	1	0	3 (3 %)	0 (0 %)	0 (0 %)	0	3 (3 %)
Between words	17	1	18 (17 %)	13	7	1	20 (19 %)	2 (2 %)	4 (4 %)	3	47 (44 %)
Between clauses	14	1	15 (14 %)	5	0	1	5 (5 %)	8 (7 %)	1 (0 %)	3	32 (30 %)
Between sentences	12	4	16 (15 %)	2	0	0	2 (2 %)	4 (4 %)	3 (3 %)	0	25 (23 %)
Total	43	6	49 (46 %)	22	8	2	30 (28 %)	14 (13 %)	8 (7 %)	6	107 (100 %)
Argumentative (N =2)											
Within words	0	0	0 (0 %)	3	0	0	3 (2 %)	0 (0 %)	0 (0 %)	0	3 (2 %)
Between words	27	2	29 (18 %)	33	3	0	36 (22 %)	14 (9 %)	10 (6 %)	9	98 (61 %)
Between clauses	9	2	11 (7 %)	1	0	1	2 (1 %)	20 (12 %)	3 (2 %)	2	38 (24 %)
Between sentences	6	2	8 (5 %)	0	0	2	2 (1 %)	12 (7 %)	0 (0 %)	0	22 (14 %)
Total	42	6	48 (30 %)	37	3	3	43 (27 %)	46 (29 %)	13 (8 %)	11	161 (100 %)

Table I-4. Number of comments for revision in stimulated recalls (high intermediate long timed group)

	Planning			Translation				No recall	Unspecified	Total
	Content	Organization	Total	Lexical retrieval	Syntactic encoding	Cohesion	Total			
Narrative (N =2)										
Below the word	0	0	0 (0 %)	2	1	0	3 (6 %)	0 (0 %)	0	3 (6 %)
At the word level	7	1	8 (15 %)	6	6	4	16 (30 %)	1 (2 %)	1	26 (48 %)
Below the clause level	4	0	4 (7 %)	4	1	1	6 (11 %)	1 (2 %)	2	13 (24 %)
At the clause level or above	0	0	0 (0 %)	1	0	0	1 (2 %)	0 (0 %)	1	2 (4 %)
At the sentence level or above	6	2	8 (15 %)	0	0	0	0 (0 %)	2 (4 %)	0	10 (19 %)
Total	17	3	20 (37 %)	13	8	5	26 (48 %)	4 (7 %)	4	54 (100 %)
Argumentative (N =2)										
Below the word	0	0	0 (0 %)	0	0	0	0 (0 %)	0 (0 %)	0	0 (0 %)
At the word level	2	0	2 (3 %)	9	2	0	11 (17 %)	2 (3 %)	1	16 (25 %)
Below the clause level	6	1	7 (11 %)	11	2	2	15 (24 %)	4 (6 %)	10	36 (57 %)
At the clause level or above	3	0	3 (5 %)	1	0	0	1 (2 %)	0 (0 %)	2	6 (10 %)
At the sentence level or above	1	1	2 (3 %)	2	0	0	2 (3 %)	1 (2 %)	0	5 (8 %)
Total	12	2	14 (22 %)	23	4	2	29 (46 %)	7 (11 %)	13	63 (100 %)

Table I-5. Number of comments for pausing in stimulated recalls (advanced short timed group)

	Planning			Translation				Monitoring	No recall	Unspecified	Total
	Content	Organization	Total	Lexical retrieval	Syntactic encoding	Cohesion	Total				
Narrative (N =2)											
Within words	0	0	0 (0 %)	1	0	0	1 (1 %)	1 (1 %)	0 (0 %)	0	2 (2 %)
Between words	24	1	25 (23 %)	28	1	3	32 (29 %)	2 (2 %)	5 (5 %)	1	65 (60 %)
Between clauses	15	3	18 (17 %)	4	1	1	6 (6 %)	2 (2 %)	1 (1 %)	0	27 (25 %)
Between sentences	4	4	8 (7 %)	0	0	1	1 (1 %)	5 (6 %)	1 (1 %)	0	15 (14 %)
Total	43	8	51 (47 %)	33	2	5	40 (37 %)	10 (9 %)	7 (6 %)	1	109 (100 %)
Argumentative (N =2)											
Within words	1	0	1 (1 %)	1	0	0	1 (1 %)	0 (0 %)	0 (0 %)	1	3 (4 %)
Between words	17	2	19 (23 %)	19	2	1	22 (26 %)	1 (1 %)	6 (7 %)	2	50 (60 %)
Between clauses	14	2	16 (19 %)	1	3	0	4 (5 %)	0 (0 %)	1 (1 %)	0	21 (25 %)
Between sentences	3	5	8 (10 %)	1	0	1	2 (2 %)	0 (0 %)	0 (0 %)	0	10 (12 %)
Total	35	9	44 (52 %)	22	5	2	29 (35 %)	1 (1 %)	7 (8 %)	3	84 (100 %)

Table I-6. Number of comments for revision in stimulated recalls (advanced short timed group)

	Planning			Translation				No recall	Unspecified	Total
	Content	Organization	Total	Lexical retrieval	Syntactic encoding	Cohesion	Total			
Narrative (*N* =2)										
Below the word	0	0	0 (0 %)	1	0	0	1 (2 %)	0 (0 %)	0	1 (2 %)
At the word level	3	0	3 (7 %)	15	1	4	20 (43 %)	4 (9 %)	1	28 (61 %)
Below the clause level	4	0	4 (9 %)	5	2	2	9 (20 %)	1 (2 %)	2	16 (35 %)
At the clause level or above	0	0	0 (0 %)	0	0	0	0 (0 %)	0 (0 %)	0	0 (0 %)
At the sentence level or above	1	0	1 (2 %)	0	0	0	0 (0 %)	0 (0 %)	0	1 (2 %)
Total	8	0	8 (17 %)	21	3	6	30 (65 %)	5 (10 %)	3	46 (100 %)
Argumentative (*N* =2)										
Below the word	0	0	0 (0 %)	4	0	0	4 (7 %)	0 (0 %)	0	4 (7 %)
At the word level	3	0	3 (5 %)	18	1	2	21 (36 %)	1 (2 %)	1	26 (45 %)
Below the clause level	1	1	2 (3 %)	7	1	1	9 (16 %)	4 (7 %)	0	15 (26 %)
At the clause level or above	2	2	4 (7 %)	2	4	0	6 (10 %)	0 (0 %)	0	10 (17 %)
At the sentence level or above	0	2	2 (3 %)	1	0	0	1 (2 %)	0 (0 %)	0	3 (5 %)
Total	6	5	11 (19 %)	32	6	3	41 (70 %)	5 (9 %)	1	58 (100 %)

Table I-7. Number of comments for pausing in stimulated recalls (advanced long timed group)

	Planning			Translation				Monitoring	No recall	Unspecified	Total
	Content	Organization	Total	Lexical retrieval	Syntactic encoding	Cohesion	Total				
Narrative (N =2)											
Within words	1	0	1 (0 %)	4	0	0	4 (3 %)	0 (0 %)	0 (0 %)	0	5 (4 %)
Between words	24	1	25 (18 %)	40	7	3	50 (37 %)	4 (3 %)	2 (1 %)	2	83 (61 %)
Between clauses	21	1	22 (16 %)	6	5	1	12 (9 %)	1 (0 %)	2 (1 %)	4	41 (30 %)
Between sentences	3	1	4 (3 %)	0	0	0	0 (0 %)	2 (1 %)	0 (0 %)	1	7 (5 %)
Total	49	3	52 (38 %)	50	12	4	66 (49 %)	7 (5 %)	4 (3 %)	7	136 (100 %)
Argumentative (N =2)											
Within words	2	0	2 (3 %)	1	1	0	2 (3 %)	0 (0 %)	2 (3 %)	0	6 (9 %)
Between words	11	0	11 (17 %)	19	2	1	22 (34 %)	2 (3 %)	0 (0 %)	0	35 (55 %)
Between clauses	7	5	12 (18 %)	2	0	0	2 (3 %)	4 (6 %)	0 (0 %)	0	18 (28 %)
Between sentences	1	3	4 (6 %)	0	0	0	0 (0 %)	1 (2 %)	0 (0 %)	0	5 (8 %)
Total	21	8	29 (45 %)	22	3	1	26 (40 %)	7 (11 %)	2 (3 %)	0	64 (100 %)

Table I-8. Number of comments for revision in stimulated recalls (advanced long timed group)

	Planning			Translation				No recall	Unspecified	Total
	Content	Organization	Total	Lexical retrieval	Syntactic encoding	Cohesion	Total			
Narrative (*N* =2)										
Below the word	0	0	0 (0 %)	3	2	0	5 (7 %)	0 (0 %)	0	5 (7 %)
At the word level	1	0	1 (1 %)	18	11	0	29 (41 %)	1 (1 %)	2	33 (47 %)
Below the clause level	4	0	4 (6 %)	9	2	1	12 (17 %)	1 (1 %)	3	20 (29 %)
At the clause level or above	1	0	1 (1 %)	4	1	0	5 (7 %)	0 (0 %)	0	6 (9 %)
At the sentence level or above	0	3	3 (4 %)	1	1	0	2 (3 %)	1 (1 %)	0	6 (9 %)
Total	6	3	9 (13 %)	35	17	1	53 (76 %)	3 (4 %)	5	70 (100 %)
Argumentative (*N* =2)										
Below the word	0	0	0 (0 %)	1	1	0	2 (3 %)	0 (0 %)	0	2 (3 %)
At the word level	1	0	1 (2 %)	16	0	4	20 (34 %)	0 (0 %)	1	22 (38 %)
Below the clause level	9	1	10 (17 %)	14	1	0	15 (26 %)	1 (2 %)	2	28 (48 %)
At the clause level or above	2	1	3 (5 %)	1	0	0	1 (2 %)	0 (0 %)	0	4 (7 %)
At the sentence level or above	2	0	2 (3 %)	0	0	0	0 (0 %)	0 (0 %)	0	2 (3 %)
Total	14	2	16 (28 %)	32	2	4	38 (66 %)	1 (2 %)	3	58 (100 %)

REFERENCES

REFERENCES

Abdel Latif, M. M. M. (2013). What do we mean by writing fluency and how can it be validly measured? *Applied Linguistics, 34*(1), 99–105.

Alexopoulou, T., Michel, M. C., Murakami, A., & Detmar, M. (2017). Task effects on linguistic complexity and accuracy: A large-scale learner corpus analysis employing Natural Language Processing techniques. *Language Learning*, *67*, 180–208.

Almond, R., Deane, P., Quinlan, T., Wagner, M., & Sydorenko, T. (2012). *A preliminary analysis of keystroke log data from a timed writing task*. (ETS Research Report No. RR-12-13). Princeton, NJ: Educational Testing Service.

Alves, R. A., Castro, S. L., & Olive, T. (2008). Execution and pauses in writing narratives: Processing time, cognitive effort and typing skill. *International Journal of Psychology*, *43*(6), 969–979.

Ädel, A. (2008). Involvement features in writing: Do time and interaction trump register awareness? In G. Gilquin, S. Papp, & M. Díez-Bedmar (Eds.). *Linking up contrastive and learner corpus research* (pp. 35–53). Amsterdam, The Netherlands: Rodopi.

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

Barkaoui, K. (2015). *Test takers' writing activities during the TOEFL iBT® writing tasks: A stimulated recall study*. (ETS Research Report No. RR-15-04). Princeton, NJ: Educational Testing Service.

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.

Barkaoui, K., & Knouzi, I. (2018). The effects of writing mode and computer ability on L2 test-takers' essay characteristics and scores. *Assessing Writing*, *36*, 19–31.

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.

Beers, S. F., & Nagy, W. E. (2011). Writing development in four genres from grades three to seven: Syntactic complexity and genre differentiation. *Reading and Writing*, *24*(2), 183–202.

Bereiter, C., & Scardamalia, M. (2009). *The psychology of written composition*. New York: Routledge.

Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge, UK: Cambridge

University Press.

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*, 5–35

Bowles, M. A. (2010). *The think-aloud controversy in second language research*. New York: Routledge.

Brown, G. T., Glasswell, K., & Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing writing*, *9*(2), 105–121.

Caudery, T. (1990). The validity of timed essay tests in the assessment of writing skills. *ELT Journal, 44*(2), 122–131.

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

Chenoweth, N. A., & Hayes, J. R. (2003). Inner voice in writing. *Written Communication, 20*, 99–118.

Chukharev-Hudilainen, E. (2014). Pauses in spontaneous written communication: A keystroke logging study. *Journal of Writing Research, 6*(1), 61–84.

Cho, Y. (2003). Assessing writing: Are we bound by only one method?. *Assessing writing*, *8*(3), 165–191.

Connor-Linton, J., & Polio, C. (2014). Comparing perspectives on L2 writing: Multiple analyses of a common corpus. *Journal of Second Language Writing*, 26, 1–9.

Deane, P. (2014). *Using writing process and product features to assess writing quality and explore how those features relate to other literacy task*s. (ETS Research Report No. RR-14-03). Princeton, NJ: Educational Testing Service.

Deane, P., Roth, A., Litz, A., Goswami, V., Steck, F., Lewis, M., Richter, T. (2018). *Behavioral differences between retyping, drafting, and editing: A writing process analysis*. (ETS Research Report No. RM-18-06). Princeton, NJ: Educational Testing Service.

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language learning*, *55*(S1), 1–25.

de Clercq, B., & Housen, A. (2017). A cross-linguistic perspective on syntactic complexity in L2 development: Syntactic elaboration and diversity. *The Modern Language Journal*, *101*(2), 315–334.

de Smet, M. J., 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.

de Smet, M. J., Leijten, M., & Van Waes, L. (2018). Exploring the process of reading during writing using eye tracking and keystroke logging. *Written Communication*, *35*(4), 411–447.

Eklundh, K. (1994). Linear and nonlinear strategies in computer-based writing. *Computers and Composition*, *11*(3), 203–216.

Eklundh, K., & Kollberg, P. (2003). Emerging discourse structure: computer-assisted episode analysis as a window to global revision in university students' writing. *Journal of Pragmatics, 35*(6), 869–891.

Elder, C., Knoch, U., & Zhang, R. (2009). Diagnosing the support needs of second language writers: does the time allowance matter? *TESOL Quarterly*, *43*(2), 351–360.

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*(1), 59–84.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College composition and communication*, *32*(4), 365–387.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, *18*(3), 299–323.

Gánem-Gutiérrez, G. A., & Gilmore, A. (2018). Tracking the real-time evolution of a writing event: Second language writers at different proficiency levels. *Language Learning*, *68*(2), 469–506.

Gass, S. M., & Mackey, A. (2017). *Stimulated Recall Methodology in Applied Linguistics and L2 Research*. New York: Routledge.

Geisler, C., & Slattery, S. (2007). Capturing the activity of digital writing: Using, analyzing, and supplementing video screen capture. In H. A. McKee & D. N. DeVoss (Eds.), *Digital writing research: Technologies, methodologies, and ethical issues* (pp. 185–200). Cresskill, NJ: Hampton Press.

Godfrey, L., Treacy, C., & Tarone, E. (2014). Change in French second language writing in study abroad and domestic contexts. *Foreign Language Annals*, *47*(1), 48–65.

Godfroid, A., & Spino, L. A. (2015). Reconceptualizing reactivity of think-alouds and eye tracking: Absence of evidence is not evidence of absence. *Language Learning*, *65*(4), 896–928.

Guo, H., Deane, P. D., van Rijn, P. W., Zhang, M., & Bennett, R. E. (2018). Modeling basic writing processes from keystroke logs. *Journal of Educational Measurement*, *55*(2), 194–216.

Hale, G. A. (1992). *Effects of amount of time allowed on the Test of Written English*. (Research Report No. 92–27). Princeton, NJ: Educational Testing Service.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp.1–28). Mahwah, NJ: Erlbaum.

Hayes, J. R. (2012). Modeling and remodeling writing. *Written communication*, *29*(3), 369–388.

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

Housen, A., De Clercq, B., Kuiken, F., & Vedder, I. (2019). Multiple approaches to complexity in second language research. *Second Language Research*. *35*(1), 3–21.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics,* 30(4), 461–473.

Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins.

Jacobs, H., Zinkgrap, S., Wormuth, D., Hartfiel, V., & Hughey, J. (1981). *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House.

Jeong, H. (2017). Narrative and expository genre effects on students, raters, and performance criteria. *Assessing Writing*, *31*, 113–125.

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.

Kellogg, R. T. (1990). Effectiveness of prewriting strategies as a function of task demands. *American Journal of Psychology*, *103*, 327–342.

Kellogg, R. T. (1994). *The psychology of writing*. New York: Oxford University Press.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp.57–71). Mahwah, NJ: Erlbaum.

Kellogg, R. T. (2001). Competition for working memory among writing processes. *The American Journal of Psychology*, *114*(2), 175–192.

Khuder, B., & Harwood, N. (2015). L2 writing in test and non-test situations: Process and product. *Journal of Writing Research*, *6*(3), 233–278.

Knoch, U., & Elder, C. (2010). Validity and fairness implications of varying time conditions on a

diagnostic test of academic English writing proficiency. *System*, *38*(1), 63–74.

Knoch, U., Rouhshad, A., & Storch, N. (2014). Does the writing of undergraduate ESL students develop after one year of study in an English-medium university?. *Assessing Writing*, *21*, 1–17.

Knoch, U., Rouhshad, A., Oon, S. P., & Storch, N. (2015). What happens to ESL students' writing after three years of study at an English medium university?. *Journal of Second Language Writing*, *28*, 39–52.

Koponen, M., & Riggenbach, H. (2000). Overview: Varying perspectives on fluency. In H. Riggenbach (Ed.) *Perspectives on fluency* (pp. 5–24). Ann Arbor: The university of Michigan press.

Kowal, I. (2014). Fluency in second language writing: A developmental perspective. *Studia Linguistica Universitatis Iagellonicae Cracoviensis*, *131*, 229–246.

Kroll, B. (1990). What does time buy? ESL student performance on home versus class compositions. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 140–154). Cambridge, UK: Cambridge University Press.

Kyle, K., & Crossley, S. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. *Language Testing*, *34*(4), 513–535.

Lambert, C., & Kormos, J. (2014). Complexity, accuracy, and fluency in task-based L2 research: Toward more developmentally based measures of second language acquisition. *Applied Linguistics, 35*(5), 607–614.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*, 590–619.

Leijten, M., & Van Waes, L. (2006). Inputlog: New perspectives on the logging of on-line writing processes in a Windows environment. In K. P. H. Sullivan & E. Lindgren (Eds.), *Computer key-stroke logging: Methods and applications* (pp. 73–93). Oxford, UK: Elsevier.

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

Leijten, M., Van Waes, L., & Ransdell, S. (2010). Correcting text production errors: Isolating the effects of writing mode from error span, input mode, and lexicality. *Written communication*, *27*(2), 189–227.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language learning*, *40*(3), 387–417.

Lindgren, E. (2005). *Writing and revising: Didactic and methodological implications of*

keystroke logging (Unpublished doctoral dissertation). Umea University, Sweden.

Lindgren, E., & Sullivan, K. P. (2003). Stimulated recall as a trigger for increasing noticing and language awareness in the L2 writing classroom: A case study of two young female writers. *Language Awareness*, *12*(3-4), 172–186.

Lindgren, E., & Sullivan, K. P. H. (2006). Writing and the analysis of revision: An overview. In K. P. H. Sullivan & E. Lindgren (Eds.), *Computer key-stroke logging: Methods and applications* (pp. 31–44). Oxford: Elsevier.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*(4), 474–496.

Lu, X. (2011). A corpus based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly, 45*, 36–62.

Malvern, D. D., Richards, B. J., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Houndmills, NH: Palgrave Macmillan.

Marian, V., Blumenfeld, H., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940–967.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, 42*(2), 381–392.

McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix.* Cambridge, UK: Cambridge University Press.

Medimorec, S., & Risko, E. F. (2016). Effects of disfluency in writing. *British Journal of Psychology*, *107*(4), 625–650.

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

New, E. (1999). Computer–aided writing in French as a foreign language: A qualitative and quantitative look at the process of revision. *The Modern Language Journal, 83*(1), 80–97

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555–578.

Oh, S. (2006). *Investigating the relationship between fluency measures and second language writing placement test decisions* (Unpublished Master's Scholarly Paper). University of Hawaii.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A

research synthesis of college-level L2 writing. *Applied Linguistics, 24*, 492–518.

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research, 31*(1), 117–134.

Papageorgiou, S., Tannenbaum, R. J., Bridgeman, B., & Cho, Y. (2015). *The association between TOEFL iBT® test scores and the Common European Framework of Reference (CEFR) levels* (Research Memorandum No. RM-15-06). Princeton, NJ: Educational Testing Service.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912.

Polio, C., & Glew, M. (1996). ESL writing assessment prompts: How students choose. *Journal of Second Language Writing*, *5*(1), 35–49.

Polio, C., & Friedman, D. A. (2017). *Understanding, evaluating, and conducting second language writing research*. New York: Routledge.

Polio, C., & Lee, J. (in press). Experimental studies in L2 classrooms. In J. W. Schwieter & Benati, A (Eds.), *The Cambridge handbook of language learning*. Cambridge: Cambridge University Press.

Polio, C., & Lim, J. (under review). Revising a writing rubric based on raters' comments: Does it result in more valid and reliable assessment? In Mehdi M. Riazi, L. Shi, & K. Barkaoui (Eds.), *An edited volume in honor of Prof. Alister Cumming*.

Porte, G. (1996). When writing fails: How academic context and past learning experiences shape revision. *System*, *24*(1), 107–116.

Powers, D. E., & Fowles, M. E. (1996). Effects of applying different time limits to a proposed GRE writing test. *Journal of Educational Measurement, 33*(4), 433–452.

Qin, W., & Uccelli, P. (2016). Same language, different functions: A cross-genre analysis of Chinese EFL learners' writing performance. *Journal of Second Language Writing, 33*, 3–17.

Quinlan, T., Loncke, M., Leijten, M., & Van Waes, L. (2012). Coordinating the cognitive processes of writing: The role of the monitor. *Written Communication*, *29*(3), 345–368.

Ranalli, J., Feng, H. H., & Chukharev-Hudilainen, E. (2018). Exploring the potential of process-tracing technologies to support assessment for learning of L2 writing. *Assessing Writing*, *36*, 77–89.

Ranalli, J., Feng, H.-H., & Chukharev-Hudilainen, E. (2019). The affordances of process-tracing technologies for supporting L2 writing instruction. *Language Learning & Technology. 23*(2), 1–11.

Révész, A., Kourtali, N. E., & Mazgutova, D. (2017). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning, 67*(1), 208–241.

Révész, A., Michel, M., & Lee, M. (2017). Investigating IELTS academic writing task 2: Relationships between cognitive writing processes, text quality, and working memory. *IELTS Research Reports Online Series*, 44.

Révész, A., Michel, M., & Lee, M. (in press). Exploring second language writers' pausing and revision behaviors: A mixed methods study. *Studies in Second Language Acquisition*.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27–57.

Roca de Larios, J., Manchón, R., Murphy, L., & Marín, J. (2008). The foreign language writer's strategic behaviour in the allocation of time to writing processes. *Journal of Second Language Writing*, *17*(1), 30–47.

Ruiz-Funes, M. (2014). Task complexity and linguistic performance in advanced college-level foreign language writing. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning: Insights from and for L2 writing* (pp. 163–192). Amsterdam: John Benjamins.

Ruiz-Funes, M. (2015). Exploring the potential of second/foreign language writing for language learning: The effects of task factors and learner variables. *Journal of Second Language Writing, 28*, 1–19.

Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of second language writing*, *9*(3), 259–291.

Sasaki, M. (2004). A multiple-data analysis of the 3.5-year development of EFL student writers. *Language Learning, 54*, 525–582.

Sasaki, M., & Hirose, K. (1996). Explanatory variables for EFL students' expository writing. *Language Learning, 46*(1), 137–174.

Schilperoord, J. (1996). *It's about time: Temporal aspects of cognitive processes in text production*. Amsterdam: Rodopi.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, *14*(4), 357–385.

Schrijver, I., Van Vaerenbergh, L., & Van Waes, L. (2012). An exploratory study of transediting in students' translation processes. *Hermes, Journal of Language and Communication in Business*, 49, 99–117.

Scott, V. M., & New, E. (1994). Computer-aided analysis of foreign language writing strategies. *CALICO Journal, 11*, 5–18.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied linguistics*, *17*(1), 38–62.

Snellings, P., Van Gelderen, A., & De Glopper, K. (2004). Validating a test of second language written lexical retrieval: A new measure of fluency in written language production. *Language Testing*, *21*(2), 174–201.

Spelman Miller, K. (2000). Academic writers online: Investigating pausing in the production of text. *Language Teaching Research, 4,* 123–148.

Spelman Miller, K. (2005). Second language writing research and pedagogy: A role for computer logging?. *Computers and Composition*, *22*(3), 297–317.

Spelman Miller, K., Lindgren, E., & Sullivan, K. P. H. (2008). The psycholinguistic dimension in second language writing: Opportunities for research and pedagogy using computer keystroke logging. *TESOL Quarterly, 42*, 433–453.

Stevenson, M., Schoonen, R., & de Glopper, K. (2006). Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. *Journal of Second Language Writing*, *15*(3), 201–233.

Tavakoli, P. (2014). Storyline complexity and syntactic complexity in writing and speaking tasks. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning: Insights from and for L2 writing* (pp. 217–236). Amsterdam: John Benjamins.

Thorson, H. (2000). Using the computer to compare foreign and native language writing processes: A statistical and case study approach. *The Modern Language Journal*, *84*(2), 155–170.

Van Hell, J. G., Verhoeven, L., & Van Beijsterveldt, L. M. (2008). Pause time patterns in writing narrative and expository texts by children and adults. *Discourse Processes, 45*(4–5), 406–427.

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.

Van Waes, L., Leijten, M., & Quinlan, T. (2010). Reading during sentence composing and error correction: A multilevel analysis of the influences of task complexity. *Reading and Writing*, *23*(7), 803–834.

Van Waes, L., & Schellens, P. J. (2003). Writing profiles: The effect of the writing mode on pausing and revision patterns of experienced writers. *Journal of pragmatics*, *35*(6), 829–853.

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.

Wallot, S., & Grabowski, J. (2013). Typewriting dynamics: What distinguishes simple from complex writing tasks?. *Ecological Psychology*, *25*(3), 267–280.

Way, D. P., Joiner, E. G., & Seaman, M. A. (2000). Writing in the secondary foreign language classroom: The effects of prompts and tasks on novice learners of French. *The Modern Language Journal, 84*(2), 171–184.

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

Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.

Wengelin, Å. (2006). Examining pauses in writing: Theories, methods, and empirical data. In K. P. H. Sullivan & E. Lindgren (Eds.), *Computer key-stroke logging and writing: Methods and application* (pp. 107–130). Oxford: Elsevier.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. Honolulu: University of Hawai'i Press.

Wu, J., & Erlam, R. (2016). The effect of timing on the quantity and quality of test-takers' writing. *New Zealand Studies in Applied Linguistics*, *22*(2), 21–34.

Wu, S. L., & Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign Language Annals*, *46*(4), 680–704.

Xu, C. (2018). Understanding online revisions in L2 writing: A computer keystroke-log perspective. *System*, *78*, 104–114.

Xu, C., & Ding, Y. (2014). An exploratory study of pauses in computer-assisted EFL writing. *Language, Learning and Technology*, 18(3), 80–96.

Yang, W. (2014). *Mapping the relationships among the cognitive complexity of independent writing tasks, L2 writing quality, and complexity, accuracy and fluency of L2 writing* (Unpublished doctoral dissertation). Georgia State University, Atlanta.

Yang, W., Lu, X., & Weigle, S. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing, 28*, 53–67.

Yoon, H. (2017). *Investigating the interactions among genre, task complexity, and proficiency in L2 writing: A comprehensive text analysis and study of learner perceptions* (Unpublished doctoral dissertation). Michigan State University.

Yoon, H., & Polio, C. (2017). The linguistic development of students of English as a second

language in two written genres. *TESOL Quarterly, 51,* 275–301.

Younkin, W. F., (1986). *Speededness as a source of test bias for non-native English speakers on the college level academic skills test* (Unpublished doctoral dissertation). University of Miami.

Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied Linguistics, 31*(2), 236–259.

Zhang, M., & Deane, P. (2015), *Process features in writing: Internal structure and incremental value over product features*. (ETS Research Report No. RR-15-27). Princeton, NJ: Educational Testing Service.