On the Effects of Disfluency in Complex Cognitive Tasks

by

Srdan Medimorec

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Psychology

Waterloo, Ontario, Canada, 2017

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Dr. Penny M. Pexman Professor
Supervisor(s)	Dr. Evan F. Risko Assistant Professor
Internal Member	Dr. Jennifer A. Stolz Professor
Internal Member	Dr. Katherine S. White Associate Professor
Internal-external Member	Dr. Catherine M. Burns Professor

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Experiments 1 to 3 have been published in the *British Journal of Psychology* (Medimorec & Risko, 2016). Experiments 4 and 5 have been published in *Cognition* (Medimorec, Young, & Risko, 2017).

Abstract

While much previous research has suggested that decreased transcription fluency has a detrimental effect on writing, there is recent evidence that decreased fluency can actually benefit cognitive processing. Across a series of experiments, I investigated the effects of experimental manipulations of transcription fluency on various aspects of essay writing (e.g., lexical sophistication), but also in the context of a single word generation task. In Chapter 1, I introduced disfluency by asking participants to type essays using one hand (vs. standard typing). The results showed that decreasing transcription fluency resulted in increased lexical sophistication. I proposed the time-based account of disfluency in composition whereby decreasing transcription fluency allows more time for lexical processes to unfold. In Chapter 2 I demonstrated that less fluent typing is not related to increased pause and revision rates. Chapter 3 provides a test between the time-based account and an account that attributes the effects to the disruption of typical finger-to-letter mappings caused by the disfluency. Here I slowed down participants' typing by introducing a delay between keystrokes. The results presented in Chapter 3 are consistent with the time-based account. In addition, in Chapter 3 I also tested the hypothesis that, unlike in previous studies, the transcription disfluency manipulation in the current study did not introduce large working memory demands. The time-based account of the effects of disfluency in composition was further supported by the results of mediation analyses presented in Chapter 4. In Chapter 5 I investigated whether effects of disfluency on lexical selection extend beyond composition to a single word generation task. I discuss implications for writing implements, and lexical selection in composition.

Acknowledgements

I am grateful to my advisor Evan Risko for tremendous support and guidance at all stages of my academic career. I want to thank Jennifer Stolz and Katherine White for their thoughtful comments, and my dedicated research assistants for spending countless hours collecting data.

Table of Contents

List of Tables

List of Figures

Introduction

Our writing tools are also working on our thoughts. --Friedrich Nietzsche

According to widely accepted models of writing (e.g., Flower & Hayes, 1981), reducing extraneous working memory load by making transcription (e.g., typing) more automatic and more fluent should allow for more efficient writing. For example, increasing typing automation should theoretically lead to more efficient writing because resources can be redistributed to higher-level processes such as planning (Bereiter & Scardamalia, 1987; Bock, 1982; Butterfield, Hacker, & Plumb, 1994; Flower & Hayes, 1981; Gentner, 1988; Hull, 1987; Kellogg, 1996; McCutchen, 1996; McCutchen, Covill, Hoyne, & Mildes, 1994; Olive & Kellogg, 2002; Pashler, 1993; Peverly, 2006). On the other hand, many individuals have speculated that making writing too fluent (or too fast) can interfere with the writing process (e.g., Heidegger, 1992; Norman, 2002). Indeed, there is recent evidence from other domains that "too much" fluency can actually impair cognitive processing (e.g., Ball, Klein, & Brewer, 2014; Grabowski, 2007; Mueller & Oppenheimer, 2014). In the present work, I sought to investigate whether decreased transcription fluency (i.e., typing speed) can benefit performance in the context of a complex cognitive task, specifically written composition, across seven experiments that included 620 participants (of those, 485 wrote essays). Overall, I analyzed more than 300,000 words of written composition.

In Chapter 1, I investigated the possibility that decreased transcription fluency can benefit some aspects of essay writing, such as lexical sophistication and syntactic complexity. Here, transcription disfluency was introduced by asking 295 participants to type essays by using only one hand (versus standard typing) across 3 experiments. Computational text analyses demonstrated that decreased fluency was related to increased lexical sophistication (i.e., increased lexical diversity and decreased word frequency), as well as to the use of fewer cohesive devices (i.e., decreased word overlap in a text). I proposed a time-based account of the effects of disfluency on lexical sophistication, whereby decreasing transcription fluency, without overly burdening working memory demands, can allow more time for lexical access to unfold, resulting in the activation of less frequent words. In the context of composition individuals may strategically select these lower frequency words and more different words, since it potentially benefits their writing. In addition, the idea that the fluency manipulation did not introduce a large increase in working memory demand was supported by subjective task load (i.e., subjective fluency) measures.

Chapter 2 includes a series of analyses that further explore the effects of disfluency on composition processes. In this chapter, I investigate the possibility that the fluency manipulation in Chapter 1 (i.e., one-handed typing) influenced lexical sophistication by affecting writing processes such as pausing and revisions. Key logged recordings from Chapter 1 were reanalyzed to obtain data about pausing and revisions in composition. The results suggest that less fluent typing is not causally related to increased pause rates (nor increased revising rate). Specifically, the manipulation of transcription fluency in Chapter 1 did not lead to more pauses or revisions.

In Chapter 3, a strong test of the time-based account is provided. Here, I contrasted the time-based account of the effects of disfluency on lexical sophistication with an account that attributes these effects to the disruption of typical finger-to-letter mappings caused by the disfluency manipulation described in Chapter 1 (i.e., one-handed typing). To do so, typing speed was decreased by introducing a slight delay between consecutive keypresses, using software developed specifically for this purpose. Critically, this manipulation did not disrupt typical finger-to-letter mappings. This study included 200 participants, and the results provided support for the time-based account of the effects of disfluency on lexical sophistication. Specifically, the results of the study demonstrated that the essays written in the less fluent condition were more lexically diverse and used less frequent words. Moreover, in this chapter I demonstrated that the disfluency manipulation (i.e., keyboard delay) did not itself introduce large working memory demands. This finding is important in explaining why the results reported here diverge from previous research that routinely reported a positive relation between transcription fluency and writing quality (e.g., Olive, Alves, & Castro, 2009).

In Chapter 4, I provide a direct test of the implied causal sequence suggested by the time-based account, whereby variation in transcription fluency between typing conditions should explain variation in lexical sophistication. The implied causal order was tested by performing a series of mediation analyses, using a large sample of 420 essays (from studies described in Chapters 1 and 3). The results of the mediation analyses demonstrated that transcription fluency mediated the relation between typing condition and the lexical sophistication measures. In addition, another set of mediation analyses demonstrated that variation in lexical sophistication across typing conditions could not be accounted for by variations in subjective fluency measures across the conditions. The results of these mediation analyses are consistent with the time-based account of the effects of disfluency on lexical sophistication.

Chapter 5 describes a set of studies aimed at investigating effects of transcription disfluency outside of the composition context. The time-based account makes specific predictions in the case of essay writing. Namely, decreasing transcription fluency allows more time for lexical processes to unfold, leading to increased lexical sophistication in composition. Given a relation between increased lexical sophistication (e.g., decreased word frequency) and better writing quality (e.g., Crossley & McNamara, 2011), there is a potential benefit in choosing lower frequency words in the context of essay writing. Thus, one potential interpretation is that individuals purposely select the activated lower frequency alternatives. A different interpretation of the effects of disfluency on writing could be that the choice of more frequent words is a result of a habitual relation between fluent typing and rapid access to frequent words (this is a variant of the disrupted access to finger-letter pairings procedural account ruled out in Chapter 3). Introducing a disfluency would presumably disrupt this relation. Such an interpretation would predict that the effects of disfluency should be observed in typing regardless of the linguistic context. I used a single word production task to investigate this notion. Across two experiments, 112 participants were asked to produce as many words as possible starting with a letter cue within a given time. While the results presented in this chapter are somewhat inconclusive, they did provide an initial indication that the effects of disfluency might exist outside of the essay-writing context.

In summary, a series of studies presented in this dissertation provides strong evidence that interfering with transcription fluency can lead to increased lexical sophistication in the context of essay writing. Specifically, slowing down individuals' typing speed resulted in decreased word frequency and increased lexical diversity. These findings provide support for the notion that specific changes to the output end of the writing process (i.e., typing) can influence writing processes by, in this case, influencing lexical selection. Implications for written composition and theories of writing are discussed.

Chapter 1

Experiments 1 to 3 have been published in the *British Journal of Psychology* (Medimorec & Risko, 2016).

The night... The night was... The night was... The night... The night was dry, yet it was raining. --Billy Crystal, Throw Momma from the Train.

Much previous research has supported the notion that increasing transcription fluency leads to improvements in writing quality (Alves, Castro, & Olive, 2008; Chenoweth & Hayes, 2001; Connelly, Campbell, MacLean, & Barnes, 2006; Connelly, Dockrell, & Barnett, 2005; Olive et al., 2009). Such results are consistent with the notion that re-distributing resources away from the "output" end of the writing process (e.g., by making transcription more automatized) can allow for more efficient writing (Fayol, 1999; Kellogg, 1996, 1999; Olive & Kellogg, 2002). However, recent evidence has suggested that "too much" fluency can actually impair cognitive processing (Mueller & Oppenheimer, 2014) and that, at least in some scenarios, introducing a disfluency can improve performance (Ball et al., 2014; Diemand-Yauman, Oppenheimer, & Vaughan, 2011). In the current study, I investigate whether decreased transcription fluency can benefit aspects of essay writing.

Transcription Fluency

As noted above, increases in transcription fluency are typically associated with increases in writing quality. For example, Olive et al. (2009) asked participants to handwrite narrative and argumentative essays using their own familiar calligraphy or an unfamiliar calligraphy (cursive uppercase). The unfamiliar calligraphy was less fluent (i.e., slower), resulted in shorter sentences, and was more effortful as measured by a secondary task response time compared to the familiar calligraphy. Moreover, essays written using the familiar calligraphy were judged to be higher quality than essays written using the unfamiliar calligraphy. Similar results have been reported by other researchers (Alves et al., 2008; Connelly et al., 2005, 2006) and the argument in most of these studies attributed the positive relation between transcription fluency and writing to the relative demands of transcription fluency on cognitive resources (Alves et al., 2008; Chenoweth & Hayes, 2001; Connelly et al., 2005, 2006; Graham, Berninger, Abbott, Abbott, & Whitaker, 1997; Kellogg, Olive, & Piolat, 2007; McCutchen, 1988; Olive et al., 2009; Wagner et al., 2011). In the current study, transcription fluency is operationalized as motor execution (i.e., typing speed) and as such should be distinguished from translation fluency, which refers to processes such as sentence generation and lexical retrieval (i.e., turning ideas into text; e.g., Peverly, 2006). Another theoretical distinction important to consider in the context of the current study is the one between disfluencies in writing and speech. While in the former case disfluency refers to a decreased transcription rate, in the latter case disfluencies refer to the use of repairs and fillers (e.g., *uh* and *oh*; Brennan & Schober, 2001; Walker, Risko, & Kingstone, 2014). Here I address the former type of disfluency.

Disfluency

While there is evidence that increased transcription fluency can benefit writing, individuals have often intuited that making writing too easy, typically with reference to typing, can impair the writing process. For example, Heidegger (1992) preferred slower handwriting over automated typewriting, stating that only handwriting was conducive to philosophical thought ("The typewriter makes everyone look the same", p. 81). Similarly, Norman (2002) presumed that handwriting encouraged "slower, more thoughtful writing" (p. 210), while more fluent typewriting hampered thinking. Indeed, Norman went further to suggest that increasing fluency further via dictating would lead to a "rambling style" (p. 210). Thus, the intuition at least is that transcription can be too fluent. Interestingly, there is some evidence to support this idea. For example, Grabowski (2007) reported that adults recalled better when writing than when speaking (i.e., the writing superiority effect). Grabowski (2007) attributed the benefits of recalling by writing to a slower pace of the recall process. More recently, Mueller and Oppenheimer (2014) reported that more fluent (or faster) laptop note taking actually impaired learning compared to less fluent (or slower) longhand note taking, even when note taking was generative (i.e., participants were asked to summarize and paraphrase lectures). The authors speculated that longhand (i.e., slower) note takers selected more important information compared to laptop note takers who engaged in verbatim note taking even when asked not to do so. Arguably, the slower pace of handwriting "forced" note takers to synthesize and summarize content unlike the faster typewriters whose typing speed enabled them to indiscriminately transcribe content. Similarly, there is evidence that handwritten essays are judged to be of better quality compared to typewritten essays (e.g., Breland, Lee, & Muraki, 2005; Bridgeman & Cooper, 1998; Neuwirth, Haas, & Hayes, 1990; Shaw, Nauman, & Burson 1994; Wolfe & Manalo, 2004).

The idea that typewriting can impair writing is clearly inconsistent with any simple linear relation between transcription fluency and writing quality. This is because typewriting is, for most, a more fluent writing mode. Relative to handwriting, the majority of students are faster at typing (Brown, 1988; Mueller & Oppenheimer, 2014). For example, handwriting speed begins to level out at about 22-24 wpm while non-expert adults typewrite 32 wpm on average (Karat, Halverson, Horn, & Karat, 1999). Similar results were reported by Alves et al. (2008) when they calculated typewriting speed using the average interval between two keystrokes within a word (I use a similar measure as an index of typewriting fluency). On average, the interval between two keystrokes within a word was 196 ms for more fluent typists, and 291 ms for less fluent typists on a dictation task, resulting in 32 and 21 wpm respectively.

Given the putative increase in transcription fluency associated with typewriting relative to handwriting why would the former be associated with lower quality writing? One potential explanation is to suggest that at some point increases in fluency fail to yield much in the way of savings in working memory but begin to inhibit writing processes that take time to operate thus leading to potentially negative effects on writing. This interpretation suggests that introducing a transcription disfluency that slows down typewriting without overly burdening working memory might potentially benefit certain processes in writing. I test this prediction in the present investigation.

Present Investigation

In the present investigation, I examine how typewriting disfluency affects different dimensions of essay writing. In a series of studies, I asked participants to typewrite narrative or argumentative essays in a standard way or by using only one hand. Thus, I interfered with transcription fluency of ostensibly skilled typists by constraining motor execution. Presumably this manipulation should not put a considerable strain on working memory resources (e.g., while the kinematics of typewriting would be changed, key location is familiar), but it should decrease transcription fluency. Matias, MacKenzie, and Buxton (1996) predicted that, with short training, one-handed typing speed on a QWERTY keyboard should be in the 61-74% range of two-handed typing speed. If we assume that average typewriting fluency is about 196 ms between keystrokes (or 32 wpm), then this would mean a reduction to about 272-341 ms between keystrokes (or 19-24 wpm) for one-handed typing which approximates average handwriting speed (note that the average typewriting fluency in the current study is ~174 ms between keystrokes in the both-handed condition, and ~318 ms between keystrokes in the one-handed condition). According to the hypothesis developed above, one-handed typewriting should potentially allow more time for activation of language processes such as lexical access without overly burdening working memory.

In the current study, I am interested in how basic processes in writing are affected by transcription fluency. To address this question, I assess the extent to which a battery of text features related to word knowledge, syntactic structure, and cohesion change as a function of the manipulation. The reliable assessment of these text features is made possible by recent advances in computational discourse processing, specifically, the development of the Coh-Metrix text analyzer (McNamara, Graesser, McCarthy, & Cai, 2014; see also the Linguistic Inquiry and Word Count (LIWC); Tausczik & Pennebaker, 2010). This approach diverges to some extent from previous research that relied, for the most part, on subjective measures of quality. While here the intuitive appeal of a general quality measure is lost (i.e., Coh-Metrix does not provide a text quality measure), a more direct and nuanced understanding of how transcription fluency influences writing is gained. In addition, in order to make contact with the previous literature I selected text features that have been demonstrated to consistently correlate with subjectively assessed writing quality (Crossley & McNamara, 2011, 2012; Crossley, Weston, McLain Sullivan, & McNamara, 2011; Guo, Crossley, & McNamara, 2013; McNamara et al., 2010, 2014).

Experiment 1

Method

Participants

Participants were 104 university students. One participant withdrew from the study, so the final sample included 103 participants (one-handed condition = 50; dominant hand = 27). All participants were compensated with course credit.

Design

I used a 2 (both-handed vs. one-handed condition) between subject design. Participants in the one-handed condition used their dominant or non-dominant hand (one hand was used throughout the task, even while revising and editing).

Stimuli and Apparatus

Participants were seated in front of a 24-in. PC monitor and asked to typewrite essays using a standard QWERTY keyboard. The essays were written in MS Word processor (versions 2010 or 2013), using the default Calibri 11pt font, with spelling and grammar check options disabled. Participants' keystroke activity was recorded using the Inputlog key logger (Leijten & Van Waes, 2013).

Procedure

Each participant wrote a timed (50 min) narrative essay (adapted from Rosenbluth & Reed, 1992; see Appendix A for prompt). Participants were asked to write a 500-word essay (i.e., one single-spaced page). Participants were informed that their essays would be graded.

Subjective Workload Measures

After completing the writing assignment, participants filled out the NASA Task Load Index (NASA-TLX, Hart & Staveland, 1988), a multidimensional workload scale (i.e., mental, physical, temporal demands, frustration, effort, and performance), ranging from -10 (i.e., low demand) to +10 (i.e., high demand; in the case of performance, -10 indicates good performance).

Measuring Linguistic Features of Essays

Essays were analyzed by using the Coh-Metrix text analyzer (Graesser, McNamara, Louwerse, & Cai, 2004; Graesser, McNamara, & Kulikowich, 2011; McNamara et al., 2014). Coh-Metrix is an automated text analyzer that computes over 100 measures of cohesion and readability indices of text. Here I focus specifically on Coh-Metrix indices indicating lexical knowledge and text difficulty (Crossley & McNamara, 2011, 2012; McNamara et al., 2010, 2014). These indices are encompassed by three broad categories: lexical sophistication, sentence complexity, and cohesion devices. Lexical sophistication is a word level category that refers to the presence of sophisticated (i.e., advanced) words in a text (Lindqvist, Gudmundson, & Bardel, 2013; McNamara et al., 2014). For example, lower frequency words are considered to be more sophisticated (Laufer & Nation, 1995). Sentence complexity indicates how difficult a syntactic construction is (Crossley & McNamara, 2011). More difficult sentence constructions include an increased number of modifiers per noun phrase, among other indices (Graesser et al., 2004). Finally, cohesive devices indicate the degree to which concepts in a text are linked (Graesser & McNamara, 2011). Thus, a more frequent use of connectives increases text cohesion (Graesser et al., 2011). More detail about individual text features is provided below.

Lexical sophistication

Lexical diversity

Lexical diversity refers to the range of vocabulary used in a text. The traditional measure of lexical diversity is type–token ratio (TTR; Templin, 1957). TTR is the ratio of unique words in the text (i.e., types) relative to the number of total words (i.e., tokens). Since TTR is highly correlated with text length (Malvern, Richards, Chipere, & Duran, 2004), a range of variants including the measure of textual lexical diversity (MTLD, McCarthy & Jarvis, 2010) and vocd - D (Malvern et al., 2004) have been developed that attempt to address this issue (by using estimation algorithms to adjust the TTR; for more details, see McCarthy & Jarvis, 2010). Greater lexical diversity adds to text difficulty (Avent & Austermann, 2003; Grela, 2002). Texts with higher lexical diversity scores are considered to have greater lexical sophistication (McNamara et al., 2014).

Word frequency

Word frequency indicates how often individual words occur in the language. Coh-Metrix computes word frequency using the CELEX database (Baayen, Piepenbrock, & Guilkers, 1996). Here I use the measures of the raw word frequency for content words and the logarithm of word frequency for all words. Word frequency is tied to text difficulty. Texts that contain less frequent words are considered more lexically sophisticated (Crossley & McNamara, 2012).

Word familiarity, word meaningfulness, and word concreteness

These word indices are based on human ratings (Coltheart, 1981; Gilhooly & Logie, 1980; Paivio, 1965; Toglia & Battig; 1978) and are tied to text difficulty. More familiar words are recognized and processed more quickly (e.g., *hell* is more familiar than *abyss*). The same is true for more meaningful words (highly meaningful words are associated with more different words; e.g., *rose* is more meaningful than *clove*), and more concrete words (e.g., *cucumber* is more concrete than *folly*). Texts that contain less familiar, less meaningful, and less concrete words are considered more lexically sophisticated (Crossley & McNamara, 2011).

Word polysemy

Word polysemy is the number of senses (i.e., meanings) a word has (e.g., McNamara et al., 2014). For example, the word *bar* has several senses: it can refer to a place where drinks are served, a court, a piece of soap, or a measure in music, among other things. Highly polysemous words can be more ambiguous and thus difficult to process (Larsen-Freeman, 2002). However, highly polysemous words are also generally more frequent (Crossley, Salsbury, & McNamara, 2010; Larsen-Freeman, 2002). Thus over time highly polysemous words become more salient and easier to process and select in appropriate contexts (Crossley et al., 2010; Larsen-Freeman, 2002).

Sentence complexity

Words before main verb and modifiers per noun phrase

Increased number of words before main verb and more modifiers per noun phrase make the syntactic structure more complex and difficult to process (Crossley et al., 2011; Perfetti, Landi, & Oakhill, 2005). Coh-Metrix calculates the mean number of words before main verb and the mean number of modifiers per noun phrase. Note that sentence length (here analyzed in the descriptive sections) can also be used as an index of sentence complexity (e.g., Medimorec, Pavlik, Olney, Graesser, & Risko, 2015).

Cohesion Devices

Aspect repetition, logical connectives, and content word overlap

Aspect repetition is an index of temporal cohesion, since it conveys information about whether an event is ongoing or completed (Duran, McCarthy, Graesser & McNamara, 2007; Klein, 1994). If these temporal signals are missing, then text difficulty increases (McNamara et al., 2014). Coh-Metrix tracks tense (e.g., present, past) and aspect (i.e., in progress or completed) across a text. When there are shifts in tense and aspect, aspect repetition decreases. Moreover, logical connectives (e.g., or, and, if-then) link the ideas in text and offer clues about text organization (Cain & Nash, 2011; Graesser, McNamara, & Louwerse, 2003). A decrease in logical connectives incidence increases text difficulty (e.g., it might create cohesion gaps; Crossley & McNamara, 2011). Finally, content word overlap indicates how frequently content words are repeated among sentences (i.e., sentences are more connected and easier to process if there is a greater overlap; McNamara et al., 2014). Decreased cohesion (i.e., less frequent aspect repetition, logical connectives, and content word overlap) increases text difficulty. The use of more cohesive devices (e.g., connectives, content word overlap) is related to lower writing quality (e.g., Crossley & McNamara, 2011), presumably because less skilled writers use cohesive devices even when they are unnecessary (e.g., McNamara, 2013). Indeed, gaps in cohesion are generally related to good writing because they introduce challenges and induce readers to actively generate more inferences (McNamara, 2013). It is also worth noting that word overlap is related to lexical diversity (i.e., more overlap indicates less diversity).

Results

A series of one-way ANOVAs were performed with condition (both-handed vs. one handed) as the factor and transcription fluency (i.e., typewriting speed), descriptive indices of text (number of words, number of sentences, average sentence length, and number of paragraphs) and the indices of lexical sophistication (TTR, MTLD, D, word frequency, word familiarity, word meaningfulness, word concreteness, word polysemy), syntactic complexity (words before main verb, modifiers per noun phrase), and cohesion (aspect repetition, logical connectives, content word overlap) as the dependent variables (note that I also compared dominant vs. non-dominant hand conditions). In addition, I report the effects of condition on various subjective workload measures (mental, physical, temporal demands, frustration, effort, performance). In the current section and throughout 95% confidence intervals are provided in square brackets [lower limit, upper limit] and Cohen's d are provided as measures of effect size when appropriate. Data cleaning and exclusion procedures are specified in the respective analyses. Where there was a violation of the homogeneity of variance an equivalent non-parametric test was also conducted. The results were qualitatively similar. Essays were not additionally edited before the analysis (results are qualitatively similar with spelling mistakes corrected). In the majority of cases, the two one-handed conditions (i.e., dominant hand vs. non-dominant hand) did not differ. When they did, I mention it in the respective analyses.

Subjective Workload Measures

This analysis includes data from 85 participants (45 in the both-handed condition) who completed the scale. Condition had a significant effect on estimated physical demand, such that it was higher in the one-handed condition compared to the both-handed condition, $F(1,83) = 39.51$, $MSE = 23.74$, $p < .001$, $d = 1.45$. The other measures were not affected, all $Fs < 2.18$, all $ps > .13$, $ds < .33$ (see Table 1).

Table 1.

Means and Standard Deviations of Essay Descriptive Measures and TLX Measures for the Two Conditions in Narrative Essays (Experiment 1)

	Both Hands		One Hand		Difference
Measure	M	SD	M	SD	Cohen's d
Essay Descriptives					
Words	640.75	158.11	598.60	179.19	.25
Sentences	30.19	8.45	29.72	11.01	.05
Words per Sentence	21.76	3.85	20.88	3.79	.23
Paragraphs	4.50	2.02	3.96	2.39	.24
TLX					
Mental Demand	-1.44	4.82	-.20	5.10	.25
Physical Demand	-7.96	2.65	-1.30	6.52	1.45^{***}
Temporal Demand	-3.02	4.83	-3.80	4.63	.16
Performance	-4.02	5.10	-3.35	5.71	.12
Effort	-1.38	4.63	.05	4.26	.32
Frustration	-5.51	5.14	-3.85	5.24	.32

*** $p < .001$ level.

Transcription Fluency

Typewriting transcription fluency was calculated as the mean keystroke interval within a word (onset of the current keypress - onset of the previous keypress in ms). I use this measure of typewriting transcription fluency because within-word keystrokes are only marginally influenced by other writing processes (Strömqvist, 1999). The keystroke intervals with a value of zero ms and the keystrokes exceeding a 2.5 SD cutoff were excluded, resulting in removal of 2.17% of keystrokes. This analysis was based on data from 101 participants (51 in the both-handed condition; no typewriting recordings were captured for two participants). Condition had a significant effect on transcription fluency, such that it was higher in the both-handed condition compared to the one-handed condition, $F(1,99) = 191.72$, $MSE = 2135.41$, $p < .001$, $d = 2.76$ (see Table 2).

Table 2.

Means and Standard Deviations of Transcription Fluency (in ms), Lexical Sophistication, Syntactic Complexity, and Cohesive Devices for the Two Conditions in Narrative Essays (Experiment 1)

	Both Hands		One Hand		Difference
Measure	M	SD	M	SD	Cohen's d
Transcription Fluency	171.04	44.11	298.38	48.26	2.76***
Type-Token Ratio	.41	.04	.45	.05	.70**
Measure of Textual Lexical Diversity	77.97	11.21	83.78	17.36	.41**
vocd-D	85.02	11.92	89.70	14.68	.35*
Log Frequency All Words	3.14	.05	3.09	.09	.72**
Word Frequency Content Words (Raw)	2.52	.11	2.45	.16	.50**
Familiarity	582.49	4.13	581.24	6.31	.24
Concreteness	349.84	18.62	352.34	18.18	.14
Meaningfulness	424.04	10.34	428.26	11.49	.39*
Polysemy	4.48	.37	4.37	.37	.28
Words Before Main Verb	4.41	1.38	3.89	1.20	.40**
Modifiers per Noun Phrase	.68	.12	.69	.13	.09
Aspect Repetition	.78	.09	.77	.09	.10
Logical Connectives	43.48	10.35	44.47	11.09	.09
Content Word Overlap Adjacent Sentences	.15	.03	.14	.04	.31
Content Word Overlap All Sentences	.13	.03	.12	.03	.31

* $p < .10$ level. ** $p < .05$. *** $p < .001$.

Descriptive Indices

There was no effect of condition on descriptive indices, all $Fs < 1.62$, $ps > .20$, $ds < .25$ (see Table 1).

Lexical Sophistication

Condition had a significant effect on lexical diversity indices such that the essays in the one-handed condition were lexically more diverse than the essays in the both-handed condition, TTR and MTLD $Fs > 4.11$, $ps < .046$, $ds > .40$, and D (marginally) $F(1,101) = 3.18$, $MSE = 177.79$, $p = .078$, $d = .35$.

Moreover, condition had a significant effect on both word frequency indices (log frequency-all words, and word frequency-content words), such that word frequency was lower in the one-handed condition, $Fs > 6.20$, $ps < .015$, $ds > .49$. Furthermore, condition affected meaningfulness (marginally) such that it was higher in the one-handed condition than in the both-handed condition, $F(1,101) = 3.84$, $MSE = 119.19$, $p = .053$, $d = .39$. The other lexical sophistication indices were not affected, all $Fs < 2.10$, $ps > .15$, $ds < .29$.

Syntactic Complexity

Condition had a significant effect on the mean number of words before the main verb such that there were fewer words before the main verb in the one-handed condition than in the both- handed condition, $F(1,101) = 4.12$, $MSE = 1.68$, $p = .045$, $d = .40$. There was no effect on the mean number of modifiers per noun phrase, $F(1,101) = .19$, $MSE = .02$, $p = .67$, $d = .09$.

Cohesion

Condition did not have a significant effect on cohesion indices, all $Fs < 2.49$, $ps > .12$, $ds < .32$ (see Table 2; note that the two content word overlap indices showed a consistent small effect, $d = .31$).

Discussion

Decreasing transcription fluency in Experiment 1 affected narrative essay writing (see Tables 1 and 2). Specifically, the essays written with one hand contained more diverse vocabulary and used less frequent words. Both of these effects have been associated with higher human judgments of essay quality. There was also some evidence that condition influenced sentence complexity, specifically, a decrease in the number of words before the main verb in the one-handed condition. However, other measures of sentence complexity (i.e., words per sentence, modifiers per noun phrase) did not show any consistent effects of the manipulation. Overall, the results of Experiment 1 provide experimental evidence that decreased fluency can benefit certain dimensions of essay writing. This result is surprising given the number of reports that interfering with transcription fluency typically negatively influences writing. As noted in the introduction, I suggest that this "benefit" of disfluency could result from the fact that typical typewriting might be too fluent and that interfering with it without introducing a large increase in working memory demand could benefit writing. The subjective workload results support this idea to some extent. There was no (or a small) effect of condition on perceived mental demand but a large effect on physical demand. Indeed, physical demand was the only workload measure to be influenced significantly. I discuss the results of Experiment 1 further following Experiments 2 and 3.

Experiment 2

Given the counterintuitive nature of the results reported in Experiment 1 I sought to replicate and extend this result in Experiment 2. Specifically, I investigate whether the effects of one-handed writing can be captured using a within-subject design, using shorter essays, a shorter time limit, and a different narrative essay prompt.

Method

Participants

Participants were university students (*N* = 75), compensated with course credit.

Design

I used a 2 (both-handed vs. one-handed condition) within subject design. Each participant was asked to write two narrative essays, one with one hand and one with both. The condition and topic order was counterbalanced.

Stimuli and Apparatus, Subjective Workload Measures, Measuring Linguistic Features of Essays

Same as in Experiment 1.

Procedure

Each participant wrote two timed (25 min), 300-word narrative essays (see Appendix B for prompts). One essay was typewritten using both hands, and the other was typewritten by using one hand (dominant or non-dominant). The rest of the procedure was the same as in Experiment 1.

Results

Analysis followed that used in Experiment 1 with the exception that within subject ANOVAs were used to assess the effect of condition. In the majority of cases, the two one-handed conditions (i.e., dominant hand vs. non-dominant hand) did not differ. When they did, I mention it in the respective analyses. There was no effect of condition order.

Subjective Workload Measures

This analysis includes data from 72 participants who completed the scales. Condition had a significant effect on estimated workload measures (i.e., mental, physical, temporal demand, effort, and frustration), such that they were higher in the one-handed condition than the both-handed condition, all *Fs* > 23.43, *ps* < .001, *ds* > .36. Moreover, participants estimated that they performed the writing task better in the both-handed condition compared to the one-handed condition, $F(1,71) = 10.22$, $MSE = 113.78$, $p = .002$, $d = .37$ (see Table 3).

Table 3.

Means and Standard Deviations of Essay Descriptive Measures and TLX Measures for the Two Conditions in Narrative Essays (Experiment 2)

	Both Hands		One Hand		Difference
Measure	M	SD	M	SD	Cohen's d
Essay Descriptives					
Words	403.72	145.30	349.99	94.66	$.45^{***}$
Sentences	19.89	9.36	17.96	7.04	$.23^{**}$
Words per Sentence	21.83	5.59	20.59	4.86	$.24^{**}$
Paragraphs	2.48	1.85	2.47	1.73	.01
TLX					
Mental Demand	-3.19	4.99	.35	4.82	$.72^{***}$
Physical Demand	-7.50	2.49	1.12	5.17	2.25^{***}
Temporal Demand	-4.26	5.10	-.82	5.45	$.65^{***}$
Performance	-5.11	4.60	-3.33	4.96	$.37^{**}$
Effort	-3.71	4.49	.50	5.03	$.88^{***}$
Frustration	-6.94	3.90	-2.87	5.47	$.87^{***}$

** p < .05 level. *** p < .001.

Transcription Fluency

Based on the established criteria I removed 1.21% of keystrokes before the analysis. Typewriting transcription fluency analysis was based on the data from 68 participants (no typewriting recordings were captured for 7 participants). Condition had a significant effect on transcription fluency, such that it was higher in the both-handed condition compared to the one-handed condition, $F(1,67) = 965.82$, $MSE = 914.67$, $p < .001$, $d = 4.12$ (see Table 4). Moreover, writing in the one-handed dominant hand condition was more fluent compared to the non-dominant condition $F(1,66) = 16.14$, $MSE = 1785.28$, $p < .001$, $d = .98$.

Table 4.

Means and Standard Deviations of Transcription Fluency (in ms), Lexical Sophistication, Syntactic Complexity, and Cohesive Devices for the Two Conditions in Narrative Essays (Experiment 2)

	Both Hands		One Hand		Difference
Measure	M	SD	M	SD	Cohen's d
Transcription Fluency	162.75	16.10	323.94	31.51	4.12***
Type-Token Ratio	.48	.05	.50	.04	.44***
Measure of Textual Lexical Diversity	79.53	16.10	78.60	16.27	.06
vocd-D	84.14	14.77	83.84	14.28	.02
Log Frequency All Words	3.12	.08	3.12	.07	.09
Word Frequency Content Words (Raw)	2.49	.14	2.48	.13	.09
Familiarity	581.84	5.51	581.60	5.12	.05
Concreteness	357.93	18.19	361.32	20.75	.17
Meaningfulness	426.78	10.30	426.56	10.87	.02
Polysemy	4.46	.43	4.40	.37	.13
Words Before Main Verb	3.98	1.21	4.28	1.50	.22**
Modifiers per Noun Phrase	.71	.15	.72	.13	.05
Aspect Repetition	.77	.08	.79	.09	.17
Logical Connectives	39.50	10.79	44.17	12.93	.39**
Content Word Overlap Adjacent Sentences	.14	.05	.14	.04	.06
Content Word Overlap All Sentences	.12	.04	.12	.04	.12

** *p* < .05 level. *** *p* < .001.

Descriptive Indices

Using all essays, condition had a significant effect on the number of words per essay, the number of sentences, and words per sentence (all indices were higher in the both-handed condition compared to the one-handed condition), all *Fs* > 4.91, *ps* < .031, *ds* > .23. There was no effect of condition on the number of paragraphs, $F(1,75) = .02$, $MSE = .41$, $p = .90$, $d = .007$ (see Table 3).

Lexical Sophistication

Condition had a significant effect on TTR, such that it was higher in the one-handed condition than the both-handed condition $F(1,74) = 16.29$, $MSE = .001$, $p < .001$, $d = .44$. The other lexical sophistication indices were not affected, all *Fs* < 1.22, *ps* > .27, *ds* < .18 (see Table 4).

Syntactic Complexity

Condition had a significant effect on the mean number of words before the main verb, such that there were more words before main verb on average in the one-handed condition than in the both-handed condition, $F(1,74) = 4.17$, $MSE = .93$, $p = .045$, $d = .22$. There was no effect of condition on the number of modifiers per noun phrase, $F(1,74) = .12$, $MSE = .01$, $p = .725$, $d = .05$.

Cohesion

Condition had a significant effect on logical connectives incidence such that it was higher in the one-handed condition than in the both-handed condition, $F(1,74) = 9.28$, $MSE = 88.22$, $p = .003$, $d = .39$. The other cohesion indices were not affected, all *Fs* < 2.65, *ps* > .10, *ds* < 18.

Summary

Clearly the results of Experiment 2 were different from those reported in Experiment 1 (compare Tables 2 and 4) in that disfluency had no systematic effects on the various text features of essays in Experiment 2.

One interesting difference between Experiments 1 and 2 (see Tables 2 and 4) is that the effect of condition on fluency (measured by Cohen's d) was much larger in Experiment 2 (d = 4.12 compared to d = 2.76 in Experiment 1).[1] As such, I compared the two experiments in order to determine the extent to which the fluency manipulation had a larger influence on transcription fluency in Experiment 2. Here, I used the fluency measures from Experiment 1 and from block one of Experiment 2. The analysis was thus based on data from 169 participants (both hands = 87). I performed a two way independent ANOVA, with Experiment (Experiment 1 vs. 2) and condition (both-handed vs. one-handed) as fixed factors, and transcription fluency as the dependent variable. Partial eta squares are reported as effect size measures. There was no effect of Experiment on fluency, $F(1,167) = 1.50$, $MSE = 1812.67$, $p = .22$, $\eta^2 = .009$. Critically, there was a significant interaction between Experiment and condition, $F(1,167) = 8.18$, MSE = 1812.67, $p = .005$, η^2 = 047. Moreover while there was no statistical difference in transcription

[1] In addition, the extent to which the within-subject design in Experiment 2 contributed to the lack of an effect of condition on the lexical indices was assessed. Here, an analysis of the Experiment 2 data restricted to the first block of trials was conducted (essentially turning the design into a between subject design). With the exception of a marginally significant effect on word polysemy (the both-handed condition > the one-handed condition, $F(1,73) = 3.60$, MSE = .55, $p = .062$, $d = .44$), condition did not affect lexical sophistication indices (all $Fs < 2$, $ps > .10$, $ds < .38$). Moreover, there was no effect of condition on syntactic complexity indices (all Fs = .30, $ps > .58$, ds = 12). Finally, condition had a significant effect only on one cohesion index (logical connectives incidence, the one-handed condition > the both-handed condition, $F(1,73)$ = 11.23, $MSE = 1569.83$, $p = .001$, $d = .77$), while the other cohesion indices were not affected, all $Fs < .76$, $ps > .39$, $ds < .21$ (similarly, there were no consistent differences between conditions in the second block of trials, nor the one-handed conditions between the two blocks). The preceding analysis suggests that the use of a within subject design was not solely responsible for the differences between Experiment 1 and Experiment 2.

fluency across the both-handed conditions between the two Experiments, $F(1,85) = 1.72$, $MSE = 1463.90$, $p = .193$, $d = .30$, there was a significant difference in the one-handed condition, such that one-handed writing was more fluent in Experiment 1 compared to Experiment 2, $F(1,87) = 6.66$, $MSE = 2183.23$, $p = .012$, $d = .59$. I discuss the importance of these results for interpreting the differences between Experiments 1 and 2 in the discussion.

Discussion

In Experiment 2 decreasing transcription fluency by having participants write with one hand had limited effects on essay writing. These results differed substantially from Experiment 1. Subsequent analyses comparing Experiment 1 and Experiment 2 suggested one potential reason. Specifically, the effect of one-handed writing seemed to have had a much larger influence on transcription fluency in Experiment 2. One potential explanation for this is that with the shorter (half the length) essays in Experiment 2 participants did not have sufficient time or practice to familiarize themselves with one-handed typewriting. For most participants, one-handed typewriting initially is unfamiliar and possibly taxing enough to limit any benefits but with increased practice (i.e., longer essays as in Experiment 1) the benefits may begin to outweigh the costs. In this sense, in Experiment 2 the short essay meant participants were made "too disfluent" for the manipulation to benefit writing processes. The subjective workload estimates seem to support the notion that longer essay writing (i.e., Experiment 1) differed substantially from shorter essay writing (i.e., Experiment 2).

The results were similar if I considered only the first block of trials (i.e., prior to exposure to the other condition). [2] Specifically, the between condition effects on the subjective effort measures were much larger in Experiment 2 than Experiment 1. Specifically, in Experiment 1 only physical demand was influenced whereas in Experiment 2 all measures were influenced. Moreover, temporal demand was influenced in different directions (i.e., participants in Experiment 1 reported less time pressure in the one-handed condition, while participants in Experiment 2 reported more time pressure in the one-handed condition). Finally, participants in Experiment 2 estimated that they performed better in the both-handed condition, while there were no differences in perceived performance in Experiment 1. This explanation, of course, is speculative. Experiment 3 attempts to put this explanation on stronger footing by returning to the longer essay format.

Experiment 3

Together Experiments 1 and 2 suggest that the potential beneficial effects of the disfluency introduced by one-handed typing might be particular to contexts that afford more of an opportunity to familiarize oneself with this novel form of transcription. Alternatively, the results of Experiment 1 could have been a type I error. Thus, in Experiment 3 I attempt to replicate and extend the results of Experiment 1 by returning to a longer essay format. I introduced one critical change in order to assess the generalizability of my results, specifically I changed the essay genre from narrative to argumentative.

Method

[2] When only the first block of trials (i.e., prior to exposure to the other condition) in Experiment 2 was considered, with the exception of performance ($F(1,70) = 2.22$, $MSE = 23.27$, $p = .14$, $d = .35$), condition had a significant effect on all measures (i.e., mental, physical, temporal demand, effort, and frustration), such that they were higher in the one-handed condition compared to the both-handed condition, all $Fs > 4.77$, $ps < .033$, $ds > .50$.

Participants

Participants were 117 university students (one hand = 59; dominant hand = 30), compensated with course credit.

Design, Stimuli and Apparatus, Subjective Workload Measures, Measuring Linguistic Features of Essays

Same as in Experiment 1.

Procedure

Each participant wrote a timed (50 min) argumentative prompt-based essay (ACT-style prompt; see Appendix C). The rest of the procedure was the same as in Experiment 1.

Results

Analysis followed that used in Experiment 1.

Subjective Workload Measures

Participants in the one-handed condition reported more physical demand, effort, and frustration, all $Fs > 4.87$, $ps < .030$, $ds > 40$. The other measures were not affected, $Fs < 2.40$, $ps > .12$, $ds < .30$ (see Table 5).

Table 5.

Means and Standard Deviations of Essay Descriptive Measures and TLX Measures for the Two Conditions in Argumentative Essays (Experiment 3)

	Both Hands		One Hand		Difference
Measure	M	SD	M	SD	Cohen's d
Essay Descriptives					
Words	643.86	213.12	578.73	132.37	.38**
Sentences	28.16	9.18	25.95	.89	.28
Words per Sentence	23.26	4.51	22.98	5.15	.06
Paragraphs	5.17	1.50	4.75	1.37	.29
TLX					
Mental Demand	-.12	3.90	.66	4.42	.19
Physical Demand	-7.53	2.61	.47	6.63	1.73***
Temporal Demand	-1.24	4.59	.14	5.03	.29
Performance	-2.91	4.35	-3.17	5.33	.05
Effort	.04	4.33	1.78	4.18	.41**
Frustration	-4.47	4.93	-1.64	5.91	.54**

** $p < .05$ level. *** $p < .001$.

Transcription Fluency

Based on the established criteria I removed 1.16% of keystrokes before the analysis. Typewriting transcription fluency analysis was based on the data from 104 participants (50 in the both-handed condition; no typewriting recordings were captured for 13 participants). Condition had a significant effect on transcription fluency, such that transcription was more fluent in the both-handed condition compared to the one-handed condition, $F(1,102) = 197.70$, $MSE = 2632.97$, $p < .001$, $d = 2.76$ (see Table 6). Note that the effect of transcription fluency reported here is comparable to that obtained in Experiment 1 and smaller than that obtained in Experiment 2. There was a difference in transcription fluency between the one-handed conditions, such that the dominant hand condition was more fluent than the non-dominant hand condition, $F(1,58) = 11.62$, $MSE = 1993.14$, $p = .001$, $d = .93$ (note that there was a consistent albeit smaller effect, $d = .30$, in Experiment 1).

Table 6.

Means and Standard Deviations of Transcription Fluency (in ms), Lexical Sophistication, Syntactic Complexity, and Cohesive Devices for the Two Conditions in Argumentative Essays (Experiment 3)

	Both Hands		One Hand		Difference
Measure	M	SD	M	SD	Cohen's d
Transcription Fluency	189.18	53.79	330.78	48.91	2.76***
Type Token Ratio	.41	.05	.43	.06	.45**
Measure of Textual Lexical Diversity	81.16	16.34	86.65	19.13	.42**
vocd-D	87.36	16.08	92.21	18.48	.28
Log Frequency All Words	3.02	.10	2.98	.08	.38**
Word Frequency Content Words (Raw)	2.34	.12	2.29	.11	.42**
Familiarity	573.80	5.18	572.50	5.30	.25
Concreteness	388.14	18.60	385.63	22.98	.12
Meaningfulness	430.28	10.12	427.69	11.03	.24
Polysemy	4.04	.32	4.00	.34	.14
Words Before Main Verb	5.58	1.96	5.47	1.38	.06
Modifiers per Noun Phrase	.81	.12	.81	.10	.04
Aspect Repetition	.74	.08	.73	.08	.06
Logical Connectives	51.48	11.78	50.71	13.14	.06
Content Word Overlap Adjacent Sentences	.13	.04	.12	.04	.39**
Content Word Overlap All Sentences	.11	.04	.10	.03	.39**

** $p < .05$ level. *** $p < .001$.

Descriptive Indices

Using all essays, condition had a significant effect on number of words, such that there were more words in the both-handed condition compared to the one-handed condition, $F(1,115) = 3.96$, $MSE = 313348.67$, $p = .049$, $d = .38$. The other descriptive indices were not affected, $Fs < 2.58$, $ps > .11$, $ds < .30$ (see Table 5).

Lexical Sophistication

Condition had a significant effect on TTR and MTLD such that the one-handed condition was more lexically diverse compared to the both-handed condition, $Fs > 5.16$, $ps < .026$, $ds > .41$, while there was no significant effect on D, $F(1,115) = 2.29$, $MSE = 300.43$, $p = .13$, $d = .28$.

Furthermore, condition had a significant effect on both word frequency measures, such that word frequency was lower in the one-handed condition than the both-handed condition, $Fs > 4.20$, $ps < .044$, $ds > .37$. The other lexical sophistication indices were not affected, $Fs < 1.81$, $ps > .18$, $ds < .25$.

Syntactic Complexity

Condition did not have a significant effect on syntactic complexity indices, $Fs < .12$, $ps > .73$, $ds < .07$. The two one-handed conditions differed statistically in the mean number of modifiers per noun phrase, such that there were more modifiers per noun phrase in the dominant hand condition than in the non-dominant hand condition, $F(1,58) = 9.55$, $MSE = .01$, $p = .003$, $d = .81$.

Cohesion

Condition had a significant effect on content word overlap between adjacent sentences, and content word overlap among all sentences, such that it was lower in the one-handed condition than in the both-handed condition *Fs* > 4.45, *ps* < .038, *ds* > .38. The other cohesion indices were not affected, *Fs* < .11, *ps* > .74, *ds* = .06 (see Table 6).

Discussion

The results of Experiment 3 replicated Experiment 1 such that decreased transcription fluency influenced various lexical indices of essays. Specifically, the essays written with one hand contained more diverse vocabulary, used less frequent words, and were less cohesive. Again, all three of these effects have been associated with higher subjectively assessed quality. The fact that all of these effects were observed again using a longer essay format supports the notion that length of essay is critical. As noted above, one potential reason length of essay might limit the "benefits" of the disfluency is that shorter essays limit the amount of practice or familiarity participants receive with one-handed typing. The explanation forwarded here for the benefit of disfluency has as one of its critical conditions that the method used to reduce transcription fluency cannot itself introduce (large) working memory demands. One-handed typing early in practice could conceivably introduce just such demands and practice should reduce them. However, it also worth noting that the results of Experiment 2 suggest that decreased transcription fluency did not have negative effects of writing (i.e., there were no systematic differences in lexical indices between the conditions). Future work directly manipulating the amount of practice (e.g., prior to writing the essay) would further elucidate the mechanisms underlying the differences I have observed between Experiments 1 and 3 and Experiment 2. Nevertheless, the observation of consistent effects across two studies while using both narrative and argumentative essay prompts suggests that disfluency can influence basic processing in writing in a manner that, at least on its face, is associated with greater essay quality.

Combined Analysis

Given the consistency of results across my longer essay studies, I next combine data from Experiments 1 and 3 to provide a test of transcription fluency effects on lexical indices on a larger sample. Thus, I analyzed narrative and argumentative essays typewritten by 220 participants (both hands = 111). I first assess the effects of condition on individual indices followed by a series of regression analyses to assess the independent effect of condition on the various text features in order to provide a more detailed analysis of the influence of disfluency on basic processes in writing. In the majority of cases there were no interactions between condition (one-handed vs. both-handed) and essay type (narrative vs. argumentative). I report significant interactions. The data are reported by fluency condition and then by essay type.

Descriptive Indices

As is apparent in Table 7 condition had a significant effect on number of words, such that there were more words in the both-handed condition compared to the one-handed condition, $F(1,218) = 5.23$, $MSE = 29992.21$, $p = .023$, $d = .32$, and number of paragraphs (marginally), such that there were more paragraphs in the both-handed condition compared to the one-handed condition, $F(1,218) = 3.71$, $MSE = 3.37$, $p = .055$, $d = .25$. There were no other effects of condition, $Fs < 1.23$, $ps > .27$, $ds < .17$.

Table 7.

Means and Standard Deviations of Essay Descriptive Measures for the Two Conditions Collapsed Across Argumentative and Narrative Essays (Combined Study)

	Both Hands		One Hand		Difference
Measure	M	SD	M	SD	Cohen's d
Essay Descriptives					
Words	642.38	188.03	587.84	155.17	.31**
Sentences	29.13	8.86	27.68	9.17	.16
Words per Sentence	22.55	4.26	22.01	4.67	.12
Paragraphs	4.85	1.79	4.39	1.95	.24*

* $p < .10$ level. ** $p < .05$.

Essay type did not have an effect on number of words, $F(1,218) = .13$, $MSE = 29992.21$, $p = .72$, $d = .05$. In contrast, essay type had an effect on number of sentences, such that there were more sentences in narrative essays compared to argumentative essays, $F(1,218) = 5.79$, $MSE = 79.68$, $p = .017$, $d = .33$. Moreover, there were more words per sentence, $F(1,218) = 9.22$, $MSE = 19.30$, $p = .003$, $d = .41$, and fewer paragraphs $F(1,218) = 8.74$, $MSE = 3.37$, $p = .003$, $d = .40$, in argumentative essays compared to narrative essays (see Table 8).

Table 8.

Means and Standard Deviations of Essay Descriptive Measures for the Two Essay Types Collapsed Across Hand Condition (Combined Study)

	Narrative		Argumentative		Difference
Measure	M	SD	M	SD	Cohen's d
Essay Descriptives					
Words	620.29	169.17	611.02	179.30	.05
Sentences	29.96	9.73	27.04	8.14	.33**
Words per Sentence	21.33	3.83	23.12	4.82	.44**
Paragraphs	4.23	2.21	4.96	1.45	.40**

** $p < .05$ level.

Lexical Sophistication

Condition had a significant effect on all three lexical diversity indices (TTR, MTLD, and D) such that they were higher in the one-handed condition compared the both-handed condition, all *Fs* > 5.10, *ps* < .026, *ds* > .30, and the two word frequency indices (log frequency-all words, and word frequency-content words) such that word frequency was lower in the one-handed condition compared to the both-handed condition, *Fs* > 11.54, *ps* = .001, *ds* > .39. Finally, there was a marginal effect of condition on familiarity, such that there were more familiar words in the both-handed than one-handed condition, $F(1,218) = 3.22$, $MSE = 27.78$, $p = .074$, $d = .22$. There were no other effects of condition, *Fs* < 2.55, *ps* > .11, *ds* < .21 (see Table 9).

Table 9.

Means and Standard Deviations of Transcription Fluency (in ms), Lexical Sophistication, Syntactic Complexity, and Cohesive Devices for the Two Conditions Collapsed Across Argumentative and Narrative Essays (Combined Study)

	Both Hands		One Hand		Difference
Measure	M	SD	M	SD	Cohen's d
Transcription Fluency	180.02	49.73	315.20	51.03	2.68***
Type Token Ratio	.41	.05	.44	.05	.55***
Measure of Textual Lexical Diversity	79.64	14.15	86.42	18.42	.42**
vocd-D	86.24	14.24	91.06	16.81	.32**
Log Frequency All Words	3.08	.10	3.03	.10	.44**
Word Frequency Content Words (Raw)	2.42	.15	2.36	.16	.40**
Familiarity	577.95	6.40	576.51	7.23	.21
Concreteness	369.85	26.69	370.36	26.67	.02
Meaningfulness	427.30	10.65	427.95	11.20	.06
Polysemy	4.25	.41	4.17	.40	.20
Words Before Main Verb	5.02	1.80	4.75	1.52	.16
Modifiers per Noun Phrase	.75	.14	.76	.13	.07
Aspect Repetition	.75	.09	.75	.09	.08
Logical Connectives	47.66	11.78	47.85	12.58	.01
Content Word Overlap Adjacent Sentences	.14	.04	.13	.04	.35**
Content Word Overlap All Sentences	.12	.03	.11	.03	.35**

** $p < .05$ level. *** $p < .001$.

Table 10.

Means and Standard Deviations of Lexical Sophistication, Syntactic Complexity, and Cohesive Devices for the Two Essay Types Collapsed Across Hand Condition (Combined Study)

	Narrative		Argumentative		Difference
Measure	M	SD	M	SD	Cohen's d
Type-Token Ratio	.43	.05	.42	.06	.17
Measure of Textual Lexical Diversity	80.79	14.74	84.94	18.11	.25*
vocd-D	87.29	13.47	89.80	17.43	.16
Log Frequency All Words	3.12	.07	3.00	.09	1.40***
Word Frequency Content Words (Raw)	2.48	.14	2.31	.12	1.33***
Familiarity	581.88	5.31	573.14	5.25	.21***
Concreteness	351.06	18.36	386.87	22.87	1.82***
Meaningfulness	426.09	11.06	428.97	10.62	.26*
Polysemy	4.43	.37	4.02	.33	1.16***
Words Before Main Verb	4.16	1.32	5.53	1.69	.91***
Modifiers per Noun Phrase	.68	.12	.81	.11	1.06***
Aspect Repetition	.77	.09	.73	.08	.44**
Logical Connectives	43.96	10.68	51.09	12.44	.62***
Content Word Overlap Adjacent Sentences	.14	.03	.13	.04	.47**
Content Word Overlap All Sentences	.12	.03	.10	.04	.53***

* $p < .10$ level. ** $p < .05$. *** $p < .001$.

There was a marginal effect of essay type on MTLD, such that it was higher in argumentative essays compared to narrative essays, $F(1,218) = 3.33$, $MSE = 267.34$, $p = .069$, $d = .25$. Moreover, essay type had an effect on both word frequency indices, familiarity and polysemy, such that they were lower in argumentative essays than narrative essays, all $Fs > 73.32$, $ps < .001$, $ds > 1.15$. In contrast, concreteness was lower in narrative essays, $F(1,218) = 179.10$, $MSE = 391.64$, $p < .001$, $d = 1.82$. The same was true for meaningfulness (marginally), $F(1,218) = 3.81$, $MSE = 115.44$, $p = .052$, $d = .26$. Moreover, there was an interaction between condition and essay type, $F(1,218) = 5.50$, $MSE = 115.44$, $p = .020$, $\eta^2 = .025$. In a simple effects analysis there was a statistical difference in meaningfulness in the both-handed condition, such that it was higher in argumentative than narrative essays, $F(1,218) = 9.34$, $MSE = 115.44$, $p = .003$, $\eta^2 = .041$, *Mean Difference* = 6.24 [2.21, 10.26], while there was no difference in meaningfulness in the one-handed condition $F(1,218) = .08$, $MSE = 115.43$, $p = .78$, $\eta^2 = .000$, *Mean Difference* = .57 [-3.50, 4.64]. The other indices were not affected, $Fs < 2.00$, $ps > .16$, $ds < .17$ (see Table 10).

Syntactic Complexity

Condition did not have a significant effect on any syntactic complexity index, $Fs < 2.31$, $ps > .13$, $ds < .16$.

In contrast, there were more verbs before main verb in narrative essays compared to argumentative essays $F(1,218) = 44.64$, $MSE = 2.32$, $p < .001$, $d = .91$, and less modifiers per noun $F(1,218) = 61.99$, $MSE = .01$, $p < .001$, $d = 1.06$.

Cohesion

Condition had a significant effect on content word overlap-all sentences and content word overlap-adjacent sentences, such that they were higher in the both-handed condition compared to the one-handed condition, *Fs* > 6,71, *ps* =.010, *ds* > .34. The other indices were not affected, *Fs* < .34, *ps* > .56, *ds* < .09.

Essay type had a significant effect on content word overlap-all sentences, content word overlap-adjacent sentences, and aspect repetition such that they were higher in narrative essays compared to argumentative essays, all *Fs* > 11.96, *ps* < .001, *ds* > .43. In contrast, there were more logical connectives in argumentative essays, $F(1,218) = 20.28$, $MSE = 136.75$, $p < .001$, $d = .62$ (see Table 10).

In summary, the results of the combined analyses demonstrate that essays written in the one-handed condition were more lexically diverse, contained more infrequent words, and were less cohesive compared to the essays written in the both-handed condition. The reported differences in log frequency-all words and MTLD between the two conditions are comparable (in absolute size) with differences between low and high proficiency writers in McNamara et al. (2010). Another way to conceptualize differences in word frequency measures between the two conditions is that the effect of disfluency is about half of that observed across genre (i.e., narrative - argumentative) in this study. Moreover, with the exception of meaningfulness, there were no interactions between condition and essay type. The later finding is a strong indicator that my fluency manipulation had a similar effect across genres despite clear changes in various indices as a function of essay genre.

Collinearity

In the next set of analyses, I explore the extent to which lexical indices representing lexical diversity, word frequency, and content word overlap are correlated. I used the combined data and one lexical index to represent lexical diversity, word frequency, and content word overlap, specifically, MTLD, log frequency-all words, and content word overlap-all sentences (results were qualitatively similar when the other indices representing lexical diversity, word frequency, and content word overlap were used). While all of the measures were correlated to some extent (see Table 11), only the correlation between MTLD and word overlap was large (i.e., > .50; Cohen, 1988), $r(218) = -.66$, $p < .001$, suggesting that the two measures are related to similar language constructs. Because condition had a larger effect on MTLD in the combined analysis, I retained that lexical index in subsequent analyses and excluded content word overlap (note that the results are qualitatively similar when using content word overlap).

Table 11.

Correlations Among Lexical Diversity, Word Frequency, and Cohesion Indices (Combined Study)

Lexical Index	1	2	3
1. Measure of Textual Lexical Diversity	-		
2. Log Frequency All Words	$-.35^{***}$	-	
3. Content Word Overlap All Sentences	$-.66^{***}$	$.23^{**}$	-

$^{**}p < .05$ level. $^{***}p < .001$.

Regression Analyses

I next used the combined data to address two questions (1) to what extent are the effects of disfluency caused by differences in number of words (i.e., the effect of condition on essay length) and (2) to what extent are the effects of disfluency on lexical diversity and word frequency independent? Both of these questions reflect the lack of independence between the measures (e.g., slowing down writing might reduce the length of essays which might impact linguistic features). I addressed these questions using a series of regression analyses. In the first step of each regression, I entered condition (both hands vs. one hand) as the IV, and one of the lexical indices (i.e., MTLD or word frequency) as the DV. In the next step, number of words was entered as an additional IV. Finally, in step 3 I entered the remaining lexical index (i.e., other than the DV) as an additional IV. In this section I report the semipartial correlations (r_s) as a measure of effect size.

In the first set of regression analyses I used Condition (0 = both hands vs. 1 = one hand) as the predictor, and MTLD as the outcome. There was a significant effect of condition on MTLD, B = 6.78 [2.42, 11.14], SE = 2.21, t(218) = 3.06, p = .002 r_s = .20, such that it was higher in the one-handed condition compared to the both-handed condition. When number of words was included in the model the effect of condition on MTLD remained significant, B = 7.02 [2.60, 11.44], SE = 2.24, t(218) = 3.13, p = .002 r_s = .21. Finally, when word frequency was entered into model the effect of condition on MTLD remained significant, B = 4.75 [.47, 9.03], SE = 2.17, t(218) = 2.17, p = .030 r_s = .14.

In the next set of regressions, word frequency was entered as the DV. There was a significant effect of condition on word frequency, B = -.04 [-.07, -.02], SE = .01, t(218) = -3.24, p = .001, r_s = -.21, such that word frequency was lower in the one-handed condition than in the both-handed condition. When number of words was added to the model, condition remained a significant predictor of word frequency, B = -.04 [-.07, -.02], SE = .01, t(218) = -3.09, p = .002, r_s = -.20. Moreover, the effect of condition on word frequency remained significant when MTLD was entered into the model, B = -.03 [-.05, -.002], SE = .01, t(218) = -2.13, p = .034, r_s = -.13.

Thus, the results of regression analyses demonstrate that condition had unique effects on both lexical diversity and word frequency. That said, the change in b values when both are in the model does suggest that to some extent the influence of condition on each variable likely influences the other (e.g., changes in word frequency might lead to changes in lexical diversity; or changes in lexical diversity might lead to changes in word frequency).

Combined Analysis Discussion

The combined results from Experiments 1 and 3 support the notion that decreased transcription fluency can influence essay writing. Again, the affected indices were lexical diversity, word frequency, and content word overlap (see Table 9). The regression analyses revealed that the effects of condition on lexical diversity (i.e., MTLD) and word frequency remained significant even when controlling for number of words. In addition, the results of regression analyses suggest that condition had independent effects on both word frequency and MTLD.

General Discussion

The current study has provided evidence that decreasing transcription fluency can benefit some aspects of writing. The results revealed that less fluent (or slower) typewriting affected both lexical sophistication and cohesion. Namely, decreased fluency was related to increased lexical sophistication (i.e., increased lexical diversity and decreased word frequency), and the use of fewer cohesive devices (i.e., decreased word overlap in a text). Both lexical sophistication and cohesion have been identified as reliable predictors of human essay judgments in previous research. Specifically, expert essay ratings increase as lexical sophistication increases and cohesion decreases (Crossley & McNamara, 2011, 2012; McNamara et al., 2014). In the following I discuss potential ways in which decreased fluency affects various lexical indices in essay writing.

Potential Mechanism Underlying Disfluency Effects in Essay Writing

In the current study I tested the prediction that decreased transcription fluency can have beneficial effects on writing. My analyses demonstrated that condition affected word frequency and lexical diversity, such that there were more infrequent words and more lexical diversity in the one-handed (i.e., less fluent) condition. One potential way to interpret these results is that decreased fluency allowed more time for lexical access, which in turn led to the activation of a larger number of words (e.g., through spreading activation; Collins & Loftus, 1975; Levelt, Roelofs, & Meyer, 1999; Roelofs, 1992). For example, there exists much evidence that individuals can access frequent words faster than infrequent words (Forster & Chambers, 1973; Knobel, Finkbeiner, & Caramazza, 2008). Thus, allowing additional time for lexical access could result in activation of less frequent words, and in the context of composition individuals may strategically select lower frequency words since this could benefit their writing. This time-dependent access of lower frequency words could also explain to some extent the increase in lexical diversity (i.e., access to more words with the increase in time). However, there was also evidence that condition influenced lexical diversity independently of its effect on word frequency. Previous research has provided evidence consistent with the idea that changes in lexical diversity do not necessarily "fall out" of using less frequent words. For example, Laufer (1994) reported that the use of less frequent words in essays did not correlate significantly with lexical diversity. Thus, lexical diversity can increase within a given "frequency band" (Laufer, 1994; Laufer & Nation, 1995; Gonzalez, 2014). Thus, the increase in lexical diversity could also be viewed as a by-product of the increased time available for lexical processes to unfold. Indeed, there is no reason that such time-dependent processes would be limited to increasing the number of low frequency words.

Conclusion

The current study represents a systematic investigation of disfluency effects in a complex cognitive task. The results support the notion that disfluency in some circumstances can benefit writing. This result, in conjunction with previous research demonstrating detrimental effects of transcription disfluencies, clearly suggests that there exist a number of rich interactions between how we write and what is ultimately written.

Chapter 2

In the following work I reanalyzed key logged data from Chapter 1 to investigate how transcription disfluency influenced composition processes, specifically pauses and revisions.

Transcription fluency (i.e., writing speed) and the number of pauses in composition are often assumed to indicate writing efficiency. For example, disfluent (or slower) typists are reported to make more pauses during composition compared to more fluent (or faster) typists (Alves, Castro, de Sousa, & Strömqvist, 2007; Deane & Quinlan, 2010; Wengelin, 2007). Such differences are interpreted to indicate different writing strategies between the two groups (Alves et al., 2007). For example, Alves et al. (2007) computationally analyzed narratives typewritten by slow and fast typists (the groups were categorized using a median split procedure) and reported that less fluent (i.e., slower) typists made a higher number of pauses during composition compared to more fluent (i.e., faster) typists. Since essays produced by the two groups were similar in lexical characteristics such as lexical density, lexical diversity, and word length, the authors concluded that less fluent typists were able to "make up" for the higher cost of transcription by making more pauses during composition, allowing them to engage in higher level writing processes (such as planning and reviewing). In Chapter 1 I reported that essays written by the less fluent group (i.e., one-handed typing) were more lexically diverse and contained less frequent words compared to the more fluent, both-handed typing group (as assessed using the Coh-Metrix; McNamara et al., 2014). Importantly, both increased lexical diversity and decreased word frequency indicate increased lexical sophistication (Crossley & McNamara, 2011, 2012). To explain these results, I proposed a time-based account, wherein decreased transcription fluency allowed additional time for lexical processes to unfold. Alternatively, if the manipulation of transcription fluency influenced pause rates, this would suggest that differences in composition could be due to processes engaged in during these pauses. Another composition process that could potentially be affected by the fluency manipulation is revisions. The process of revision could potentially influence text characteristics (e.g., lexical sophistication) through operations that include editing errors (e.g., spelling) and transforming the text (e.g., deleting parts of the text; Butler, & Britt, 2011; Piolat, Roussey, Olive, & Amada, 2004; Stevenson, Schoonen, & de Glopper, 2006). Thus, I also investigate the causal relation between transcription fluency and revisions in composition.

Present Investigation

In the current study I used a large set of narrative and argumentative typewritten essays (from Chapter 1) to analyze the influence of an experimental manipulation of transcription fluency on pauses and revisions in composition. Here, I report a pause rate measure, representing an average number of pauses per text boundary (i.e., pause rate between words, sentences, and paragraphs) using a pause threshold of 2 seconds (a commonly used threshold; e.g., Alves et al., 2007; Levy & Ransdell, 1995; Medimorec & Risko, 2017; Strömqvist & Ahlsén, 1999). I also report the number of revisions (i.e., the total number of revisions during composition).

Method

Participants

Participants were 204 university students (both-handed typing = 101; Chapter 1).

Stimuli and Apparatus, Design, Procedure

Same as in Chapter 1 (Experiments 1 and 3).

Measuring Pauses

Pause threshold was set to 2 s (2 s included). Pauses were captured by the Inputlog key logging software (Leijten & Van Waes, 2013). Inputlog identifies pauses at different text locations by using an algorithm (see Leijten & Van Waes, 2013). In the current study, pauses were analyzed at three locations within a text: between words, sentences, and paragraphs.

For example, pauses between words represent a sum of pauses between the onset of the last letter of a given word and onset of the first letter of the following word (e.g., pauses after words are latencies between the last letter of the previous word and the spacebar, while the pauses before words are latencies between the spacebar and the first letter of the current word). The reported pause rates are frequencies per lexical unit (e.g., the rate between words is calculated as the overall pause count at word boundaries/number of words). Revisions were also determined using Inputlog. I used all detected revisions.

Results

All statistical analyses were carried out on log10 transformed pause data to address positive skewness. The results were qualitatively similar when raw data were used.

Pause Rates at Different Text Boundaries

I performed a series of univariate analyses, with condition (one-handed vs. both-handed typing) and essay type (narrative vs. argumentative) as fixed factors, and pause rate at different text boundaries (i.e., between words, sentences, and paragraphs) as the dependent variable. Cohen d's are reported as effect sizes. There were no significant interactions between condition and essay type at any text boundary.

There were no significant effects of condition (one-handed vs. both-handed typing) on pause rates between words, sentences, or paragraphs, all $Fs < .91$, $ps > .34$, $ds < .15$ (mean values and standard deviations for pause rates across the conditions are presented in Table 12).

Table 12.

Mean Pause Rates, Standard Deviations, and Effects Sizes (Cohen's d) by Condition, Log Transformed Data

Pause Location	Condition				
	Both Hands		One Hand		Effect Size
	M	SD	M	SD	d
Words	-.229	.041	-.223	.044	.14
Sentences	-.203	.092	-.195	.094	.09
Paragraphs	-.140	.156	-.148	.156	.05

Essay type affected pause rates between words such that there were more pauses in argumentative essays compared to narratives, $F(1,200) = 8.99$, $MSE = .002$, $p = .003$, $d = .43$. There were no effects of essay type on pause rates between sentences, or paragraphs, *Fs* < .14, *ps* > .71, *ds* < .06, $BF_{01}s > 6.14$ (see Table 13).

Table 13.

Mean Pause Rates, Standard Deviations, and Effects Sizes (Cohen's d) by Essay Type, Log Transformed Data

Pause Location	Essay Type				
	Narrative		Argumentative		Effect Size
	M	SD	M	SD	d
Words	-.235	.038	-.217	.046	.43**
Sentences	-.201	.096	-.196	.091	.05
Paragraphs	-.136	.156	-.151	.155	.10

** $p < .05$ level.

To assess the strength of the observed null effects, I performed a series of Bayesian ANOVAs (using the JASP program; Love et al., 2015), with condition and essay type as fixed factors, and pause rates at each text boundary as the dependent variable. When condition was the fixed factor, the results provided positive evidence (i.e., $BF_{01} > 3$; Kass & Raftery, 1995) in support of H0 in each case, all $BF_{01}s > 4.13$. The results were similar when essay type was the fixed factor (at sentence and paragraph levels), $BF_{01}s > 4.99$.

In summary, there was no effect of a manipulation of transcription fluency on pause rates between words, sentences, or paragraphs. In contrast, essay type affected pause rates between words such that there were more pauses in argumentative than narrative essays (pause rates between sentences and paragraphs were not affected).

Revisions

I performed a univariate analysis, with condition (one-handed vs. both-handed typing) and essay type (narrative vs. argumentative) as fixed factors, and the number of revisions as the dependent variable. There was no significant effect of condition on the number of revisions, $F(1,201) = 1.57$, $MSE = 112334.35$, $p = .21$, $d = .08$, $BF_{01} = 4.00$ (BH: $M = 466.94$, $SD = 358.59$; OH: $M = 496.90$, $SD = 357.61$). Essay type affected the number of revisions such that there were more revisions in argumentative essays compared to narratives, $F(1,201) = 17.43$, $MSE = 112334.35$, $p < .001$, $d = 1.29$ (Narrative: $M = 242.72$, $SD = 155.89$; Argumentative: $M = 603.93$, $SD = 404.98$).

Thus, similar to pausing, there was no effect of the manipulation of transcription fluency on revisions, while essay type affected revising such that there were more revisions in argumentative than narrative essays.

Discussion

Reanalysis of the Experiments in Chapter 1 demonstrated that less fluent typing is not causally related to increased pause rates nor increased revising rate. This is not because of any lack of an ability to detect an effect of the disfluency manipulation on pause rates and revisions given the demonstration that individuals paused more at the word boundary when composing argumentative than narrative essays and also revised more in argumentative essays. Both of the latter effects likely reflect the greater cognitive effort associated with argumentative text composition (e.g. van Hell, Verhoeven, & van Beijsterveldt, 2008). Thus the results of the current study suggest that the reported differences in composition as a function of fluency are not due to differences in pauses and revisions across the conditions. Rather, the results are consistent with the interpretation that decreased transcription fluency allowed additional time for lexical processes (e.g., lexical access) to unfold (i.e., the time-based account).

One potential way to explain the lack of a relation between disfluency and pause rates is that both increased pause rates and decreased transcription fluency observed in previous studies (e.g., Alves et al., 2007; Medimorec & Risko, 2017) are caused by the cognitive demands associated with composition. For example, if composition is cognitively demanding for a given typist this might slow transcription speed and increase pause rates. The same is true for revisions.

Conclusion

The current study systematically investigated the relation between decreased transcription fluency, and pausing and revising during composition. The results support the notion that transcription fluency and pause and revision rates are not causally related.

Chapter 3

The following work has been published in *Cognition* (Medimorec, Young, & Risko, 2017).

Changes have been introduced to improve the flow of the dissertation.

That's not writing, that's typing. --Truman Capote

Recent research has suggested that in some circumstances, introducing disfluency can benefit performance (e.g., Ball et al., 2014; Mueller & Oppenheimer, 2014). A surprising example of this general phenomenon was reported in Chapter 1, where decreasing transcription fluency (or typing speed) by having individuals type with one hand resulted in more lexically sophisticated essays. This finding is surprising theoretically because transcription fluency is typically thought to be positively correlated with writing quality as the automatization of transcription arguably affords the re-distribution of resources to higher level writing processes such as planning (Fayol, 1999; Kellogg, 1999). Thus understanding how transcription disfluency influences aspects of writing provides a unique opportunity to gain a deeper understanding of the basic mechanisms underlying lexical selection in written composition (and language use in general). Here I provide a strong test between two potential theoretical accounts of the effect of disfluency on lexical sophistication.

Typing disfluency and cognition

In Chapter 1, transcription disfluency was introduced by having participants typewrite essays using one hand (vs. standard typing). When essays were computationally analyzed, the results demonstrated that essays typewritten in the less fluent (or slower) condition were more lexically sophisticated (i.e., they exhibited increased lexical diversity and decreased word frequency). I interpreted these results as consistent with the idea that typewriting may be too fluent (Heidegger, 1992; Norman 2002). For example, Norman (2002) suggested that handwriting encouraged more thoughtful writing compared to typewriting because the former was slower. I proposed a time-based account of the effects of disfluency on lexical sophistication whereby decreasing transcription fluency allowed more time for lexical processes, such as lexical access, to unfold, ultimately allowing individuals to strategically select lower frequency alternatives since this could benefit their writing. Such a mechanism is consistent with the general underpinnings of most theories of lexical processing, which posit that accessing infrequent words takes more time than accessing more frequent words, both in speech and writing (Forster & Davis, 1984; Navarrete, Basagni, Alario, & Costa, 2006). For example, Crowe (1998) used a verbal fluency task to investigate lexical selection and found that participants first generated the more available frequent words followed by infrequent words. According to the time-based disfluency account, the critical variable in producing this effect on lexical sophistication should be the delay in transcription speed caused by disfluency.

While writing with one hand (relative to two) certainly slows down transcription, it also interferes with writing in other ways. For example, relatively skilled typewriting involves specific pairings of fingers and keyboard keys (Purcell, Napoliello, & Eden, 2011), and switching to one-handed typing would disrupt those mappings. This disruption could in theory influence lexical selection during writing. For example, there is evidence that individuals exhibit different letter preferences when typing on a QWERTY keyboard across different input modalities depending on whether they use both hands or only one hand (Pelleg, Yom-Tov, & Gabrilovich, 2015). Thus, the way that we type can influence what we type. An effect on lexical sophistication would occur if lexical selection was influenced by the frequency of motor production (i.e., a bias towards selecting often typed words) and one-handed typing interfered with such an effect. In other words, the decreased lexical sophistication in the more fluent (or standard) typing condition could potentially be explained by a habitual combination of more rapid word access and the more automatized typing of those words (compared to infrequent words), with the latter being disrupted when we type one-handed.

In order to test between the two accounts described above, it is necessary to find a manipulation that could slow down regular (both-handed) typing without disrupting the finger-to-letter mappings. To this end, I had software developed that allowed me to introduce a delay between keypresses. Thus, I could, relatively directly, control typing speed while individuals used their familiar two-handed typing (and thus maintained the same finger-to-letter relations across conditions). To my knowledge, this is the first time that the effects of such a manipulation have been reported.

In addition to the two accounts described above, I also examined the idea that fluency effects in composition could result from effects on participants' subjective task experience. Previous work has provided evidence that conscious experience of low effort or high speed, referred to as subjective fluency (Winkielman, Schwarz, Fazendeiro, & Reber, 2003), can influence cognitive processing. For example, subjective fluency can elicit a positive affective reaction (Schwarz & Clore, 1996) and facilitate creativity (Nadler, Rabi, & Minda, 2010). Moreover, individuals can perceive the same objective experiences as more or less fluent depending on previous experience and expectations (Whittlesea, 1993). Thus, many fluency effects can involve a subjective fluency component. To test this notion, participants' responses to the NASA Task Load Index (Hart & Staveland, 1988) were examined. Finally, since increased lexical sophistication represents one of the predictors of better essay quality (Crossley & McNamara, 2011), I investigate whether disfluency effects on lexical sophistication extend to human essay quality judgements.

Experiment 4

Method

Participants

A total of 202 university students participated, but two participants did not complete the study. Sample size was determined using the effect size of .40 (based on Chapter 1), and power of .80 (Faul, Erdfelder, Buchner, & Lang, 2009).

Design

A 2 (standard vs. keyboard delay condition) between-subject design was used.

Stimuli and apparatus

The essays were written using a QWERTY keyboard and software that controlled the minimum time delay between keystrokes. Based on piloting, I set this minimum delay to 100 ms.[3] Keystroke activity was recorded using the delay application and the Inputlog key-logger (Leijten & Van Waes, 2013).

Procedure

Participants wrote a timed (50 min) argumentative essay (used in Chapter 1). Participants were asked to write at least 500 words and were informed that their essays would be graded. Participants were given a 3-minute practice session to get familiarized with the keyboard by typing a sentence. After the writing task, participants filled out the NASA-TLX scale.

Measuring transcription fluency and linguistic features of essays

Transcription fluency was determined by calculating times between consecutive lower case letters recorded by the delay application. I removed 1.42% of keystrokes within individual participants exceeding 2.5 SD of the mean (for 199 participants; one participant used all caps).

[3] Note that this does not necessarily lead to a 100 ms decrease in typing speed because the majority of keypresses in regular typing already exceed 100 ms (~73% in the standard condition). Thus, the introduced delay could be roughly conceptualized as a removal of all sub 100 ms interword keypresses from standard typing.

Essays were analyzed using the Coh-Metrix text analyzer (McNamara, Graesser, McCarthy, & Cai, 2014). I also report corresponding language indices using the Tool for the Automatic Analysis of Lexical Sophistication (TAALES; Kyle & Crossley, 2015), and the Tool for the Automatic Analysis of Cohesion (TAACO; Crossley, Kyle, & McNamara, 2015). I focus on two relatively independent lexical sophistication indices affected by the transcription fluency manipulation in Medimorec and Risko (2016), specifically, lexical diversity and word frequency.

Lexical diversity

Lexical diversity measures (the type–token ratio (TTR); the measure of textual lexical diversity (MTLD), and vocd-D) were derived from Coh-Metrix. TTR results were also derived from TAACO.

Word frequency

Word frequencies from the CELEX database (the log-frequency for all words and the raw word frequency for content words) are derived by Coh-Metrix. Word frequencies from the SUBTLEXus (Brysbaert & New, 2009) and British National Corpus (BNC; BNC Consortium, 2007) databases are derived by TAALES.

Essay grading

Three trained raters independently and blind to condition scored essays using a 6-point rating scale (based on the ACT Writing Test Scoring Rubric). Raters were trained to use the rubric by scoring two samples of argumentative essays from another corpus ($N = 30$) until the averaged interrater reliability reached $r > .50$. When the essays from the current study were scored, the interrater reliability was significant, $r(198) = .56$, $p < .001$. The mean score between the two raters with the highest correlation was used as the final score of each essay. If the differences between these two raters were ≥ 1, then the score closest to the third rater's score was used.

Results

A series of one-way ANOVAs was performed with condition (standard vs. keyboard delay) as the factor and transcription fluency, lexical sophistication, subjective fluency, and essay quality as the dependent variables. Essays were not edited before the analysis (results are qualitatively similar with spelling mistakes corrected).

Descriptive essay indices

Condition affected transcription fluency such that it was more fluent in the standard condition, $F(1,197) = 110.14$, $MSE = 2796.13$, $p < .001$, $d = 1.50$. While there were more words typed in the standard condition, $F(1,198) = 6.35$, $MSE = 14329.49$, $p = .013$, $d = .36$, essays in the delay condition contained longer words (letters and syllables per word), $Fs > 4.17$, $ps < .043$, $ds > .28$. There were no other differences (see Table 14).

Table 14.

Means, Standard Deviations, and Cohen's d (Cohen, 1988) of Essay Descriptive Measures for the Two Conditions (Experiment 4)

	Standard		Delay		Difference
Measure	M	SD	M	SD	Cohen's d
Transcription Fluency	223.37	58.47	302.05	46.55	1.50***
Number of Words	593.66	138.56	551.01	97.27	.36**
Word Length (letters)	4.68	.24	4.75	.25	.29**
Word Length (syllables)	1.49	.08	1.51	.09	.31**
Sentence Count	25.91	6.48	25.39	6.55	.08
Words per Sentence	23.57	4.90	22.55	4.92	.21
Paragraph Count	4.12	1.71	4.15	1.72	.02

** *p* < .05 level. *** *p* < .001.

Lexical sophistication

Condition had a significant effect on type-token ratio, $F(1,198) = 9.70$, $MSE = .002$, $p = .002$, $d = .44$, and vocd-D (marginally), $F(1,198) = 3.29$, $MSE = 264.62$, $p = .071$, $d = .26$, such that they were higher in the delay condition. There was no effect on the measure of textual lexical diversity, $F(1,198) = 2.13$, $MSE = 239.96$, $p = .146$, $d = .21$, though the pattern of means was in the same direction. Moreover, condition had an effect on both word frequency indices such that they were lower in the delay condition, log word frequency-all words, $F(1,198) = 4.49$, $MSE = .01$, $p = .035$, $d = .30$, raw word frequency-content words, $F(1,198) = 4.74$, $MSE = .02$, $p = .031$, $d = .31$ (see Table 15). The results were similar when lexical diversity and word frequency from TAACO and TAALES were used (see Table 16).

Table 15.

Means, Standard Deviations, and Cohen's d of Lexical Diversity and Word Frequency for the Two Conditions, Coh-Metrix Indices (Experiment 4)

	Standard		Delay		Difference
Measure	M	SD	M	SD	Cohen's d
Type-Token Ratio	.41	.04	.43	.05	.44**
Measure of Textual Lexical Diversity	80.40	14.43	83.60	16.49	.21
vocd-D	87.06	14.81	91.23	17.61	.26*
Log Frequency All Words	3.03	.09	3.01	.08	.30**
Word Frequency Content Words (Raw)	2.37	.12	2.33	.12	.31**

* $p < .10$ level. ** $p < .05$.

Table 16.

Means, Standard Deviations, and Cohen's d of Lexical Diversity and Word Frequency for the Two Conditions, TAACO and TAALES Indices (Experiment 4)

	Standard		Delay		Difference
Measure	M	SD	M	SD	Cohen's d
Type-Token Ratio	.42	.04	.44	.05	.46**
Log Frequency All Words (SUBTLEXus)	4.44	.14	4.40	.13	.35**
Log Frequency All Words (BNC)	4.91	.08	4.87	.09	.37**
Log Frequency Content Words (SUBTLEXus)	3.77	.17	3.71	.18	.35**
Log Frequency Content Words (BNC)	4.26	.10	4.22	.12	.34**

** $p < .05$ level.

Subjective fluency

Condition had an effect on two of the subjective fluency measures, namely physical demand, $F(1,198) = 12.34$, $MSE = 30.20$, $p = .001$, $d = .50$, and frustration, $F(1,198) = 27.46$, $MSE = 32.12$, $p < .001$, $d = .74$, such that they were higher in the delay condition. The other measures were unaffected, *Fs* < 2.27, *ps* > .13, *ds* < .22 (see Table 17).

Table 17.

Means, Standard Deviations, and Cohen's d of Subjective Fluency Measures for the Two Conditions (Experiment 4)

	Standard		Delay		Difference
Measure	M	SD	M	SD	Cohen's d
Mental Demand	.16	4.80	1.10	4.71	.20
Physical Demand	-5.27	4.81	-2.54	6.10	.50**
Temporal Demand	-1.82	5.05	-1.74	5.43	.02
Performance	2.26	4.51	1.31	4.96	.20
Effort	1.12	4.13	1.98	3.96	.21
Frustration	-2.78	5.48	1.42	5.85	.74***

** $p < .05$ level. *** $p < .001$.

Essay quality

There was no effect of condition on human-judged essay quality, $F(1,198) = .55$, $MSE = .19$, $p = .458$, $d = .11$. Graders rated the essays written with the standard keyboard ($M = 4.24$ SD = .42) as similar to those written with the delayed keyboard ($M = 4.19$ SD = .45). Grades correlated positively with the number of words, $r(198) = .41$, $p < .001$. Controlling for the number of words, lower word frequency was related to higher quality, raw word frequency-content words B = -.84, SE = .22, $t(197)$ = -3.84, p < .001, r_s = -.24, log word frequency-all words, B = -1.16, SE = .31, $t(197)$ = -3.79, p < .001, r_s = -.24, whereas lexical diversity measures were not related to quality judgements, Bs < .58, ts < .81, ps < .44, $r_s s$ < .06.

Regression analysis

I performed a regression analysis to determine the extent to which the effects of disfluency on lexical indices are related to differences in the number of words and subjective fluency (since condition affected the number of words, physical demand, and frustration). I entered condition (0 = standard, 1 = delay) and the number of words as the independent variables, and raw frequency-content words or type-token ratio (indices with the largest effect sizes) as the dependent variable. In the next step, I entered physical demand and frustration as additional independent variables. Effect sizes are semipartial correlations (r_s).

Condition had a significant effect on raw frequency-content words, B = -.040, SE = .02, $t(197)$ = -2.27, p = .025, r_s = -.16, and type-token ratio, B = .012, SE = .01, $t(197)$ = 2.13, p = .034, r_s = .13, controlling for number of words. The effect remained significant with physical demand and frustration in the model for raw frequency-content words, B = -.044, SE = .02, $t(195)$ = -2.36, p = .019, r_s = -.17, and type-token ratio (marginally), B = .010, SE = .01, $t(195)$ = 1.74, p = .083, r_s = .11.

Discussion

The results of the current study provide clear evidence against the mapping disruption account of disfluency on lexical sophistication. In particular, by introducing a delay between consecutive key presses, I was able to slow typing without disrupting familiar finger-to-letter mappings. Despite the latter, I still observed a significant effect of disfluency on lexical sophistication. Specifically, the essays written in the keyboard delay condition were more lexically diverse and used less frequent words. This result is consistent with the time-based account suggested in Chapter 1 wherein slowing typing provides additional time for lower frequency alternatives and more alternatives to be activated. Presumably in the context of essay writing individuals strategically select lower frequency alternatives since this could improve their writing.

The transcription fluency manipulation also influenced subjective estimates of physical demand and frustration. Interestingly, the effect of disfluency on lexical diversity (marginally) and word frequency remained after controlling for these effects. Thus, while introducing disfluency in transcription clearly has a marked effect on composers' subjective experiences, there appears to be an effect of disfluency on lexical sophistication that is independent of these effects. That said, the hypothesis that subjective fluency can affect lexical sophistication remains interesting to further pursue given that I only examined a limited number of subjective measures. In addition, I have focused on a specific set of lexical indices tied to lexical sophistication and there exists a large number of other lexical measures that could be influenced by introducing transcription disfluency and the resulting subjective experiences.

Lastly, the observed differences in lexical sophistication did not translate to an effect on subjective essay quality. There are a number of potential reasons for this result. First, it is possible that the measure of grading was not sufficiently sensitive. For example, the reliability between essay graders was not particularly high. Second, it is possible that while lexical sophistication is related to essay quality, the magnitude of the effect of transcription disfluency on lexical sophistication (here $d \sim .30$) is not large enough to translate into human-graded essay quality. Lastly, it is also possible that the previously reported relation between lexical sophistication and essay quality (e.g., Crossley & McNamara, 2011) is confounded by some third variable. For example, rather than the use of lower frequency words causing higher essay scores, it might be that individuals who tend to use lower frequency words are superior on other aspects of writing that are more directly related to subjective essay quality. In the latter case, a manipulation that influences lexical sophistication but not these other aspects of writing would not be expected to translate into an effect on subjective essay quality.

The disfluency effects in written composition reported here diverge to some extent from previous research that reported a positive relation between transcription fluency and writing quality. For example, when participants in Olive et al. (2009) wrote essays using a disfluent cursive uppercase calligraphy, they produced essays that were judged to be of lesser quality than those written in participants' own familiar calligraphy. Similarly, Alves et al. (2008) reported that slower typists produced texts that contained fewer different words compared to faster typists. Why the discrepancy? As I argued in Chapter 1, unlike previous studies, the transcription disfluency manipulation in the current study did not introduce large working memory demands. Consistent with this notion, there was no significant effect of condition on perceived mental demand or performance (I provide a more explicit test of this notion in Experiment 5). Not burdening working memory is important because lexical selection is likely tied to the availability of working memory. Indeed, previous research has demonstrated that increasing working memory load decreases verbal fluency (Rosen & Engle, 1997).

Experiment 5

The results presented in Chapter 1 and in the current chapter demonstrated that the essays written in disfluent (i.e., one-handed typing, keyboard delay) conditions were more lexically diverse and used less frequent words. Together, these findings are consistent with the notion that slowing typing could lead to increased lexical sophistication. According to the proposed time-based account of the effects of disfluency on lexical sophistication, decreasing transcription fluency should allow more time for lexical processes, such as lexical access, to unfold. As noted previously, the disfluency effects in written composition reported in Chapters 1 and 3 differ from previous research that typically reported a positive relation between transcription fluency and writing quality. I suggested that the reported benefit of disfluency resulted from the fact that the disfluency manipulation introduced here did not itself increase working memory demands, unlike fluency manipulations in previous studies (e.g., Olive et al., 2009). While such an explanation was supported by subjective workload measures in Chapters 1 and 3, in Experiment 5 I seek to provide a stronger test of the notion that the disfluency I introduced does not necessarily burden working memory resources.

Working Memory Demands in Composition

Writing is a cognitively taxing process (Flower & Hayes, 1981; Graham, Gillespie, & McKeown, 2013; Kellogg, 1994; Kellogg & Whiteford, 2009; McCutchen, 1988). For example, in Flower and Hayes' (1981) cognitive model of writing there are different recursive and interleaved components (e.g., planning, transcription) all of which place heavy and competing demands on working memory resources (but also see Kellogg, 2001; McCutchen, 2000; Olive, 2004; Olive at al., 2009). Such a model of writing implies that increasing transcription automation should lead to better writing because additional working memory resources can be redistributed to higher level processes such as planning (e.g., Flower & Hayes, 1981; Kellogg, 1996; Peverly, 2006). Thus, increasing demands on transcription should interfere with other writing processes. For example, when Olive et al. (2009) manipulated fluency by asking participants to write essays using cursive uppercase (versus more fluent familiar calligraphy) the essays written in the disfluent condition were judged to be of lower quality, presumably a consequence of increased working memory demands introduced by the disfluency (e.g., Alves et al., 2008; Kellogg et al., 2007; McCutchen, 1988; Olive et al., 2009; Wagner et al., 2011). The results in Chapters 1 and 3, however, appear inconsistent with this prediction. Impairing transcription led to more lexically sophisticated writing, arguably because the disfluencies I introduced slowed down transcription without adding additional demands on working memory.

In the current experiment, I set out to investigate whether the disfluency manipulation introduced in Experiment 4 (i.e., keyboard delay) led to an increase in working memory demands compared to standard typing. Working memory demands during writing can be reliably assessed by using a dual-task technique (e.g., Olive & Kellogg, 2002). In this task, participants are asked to execute an additional task during composition.

Specifically, in the current study I asked participants to detect auditory probes (i.e., the secondary task) as quickly as possible during composition (i.e., the primary task; modeled after Olive & Kellogg, 2002). Participants first performed only the secondary task (i.e., the baseline single task condition), followed by three dual task conditions (standard typing, and two delay keyboard conditions). Performance in the secondary task is expected to decrease with increased primary task demands. Since different writing processes (e.g., planning, reviewing, transcription) share and compete for working memory resources (e.g., Kellogg, 2001), any potential differences in transcription demands between standard and disfluent conditions should be detected by the secondary task reaction time (i.e., primary and secondary tasks compete for limited resources). In other words, the secondary task should measure the amount of available spare capacity. Thus, if the disfluency manipulation indeed does not introduce additional working memory demands, then we should expect no differences in reaction times to auditory probes between standard and disfluent typing conditions.

Method

Participants

Participants were 33 university students.

Design

A 4 (baseline, and keyboard delay: 0, 100, or 300 ms) within subject design was used. The order of conditions was counterbalanced.

Stimuli and Apparatus

The essays were written using a QWERTY keyboard and software that controlled the minimum time delay between keystrokes (the same software was used in Experiment 4). Auditory probes (i.e., tones) were presented using external speakers and participants responded by pressing an affixed mouse button placed between them and the keyboard.

Procedure

In the first part of the experiment, participants performed only the response time (RT) task in order to compute their mean baseline RTs. Participants were asked to react as quickly as possible whenever they detected an auditory probe by pressing the button in front of the keyboard. Auditory probes appeared on average every 10 seconds (5, 10, or 15 s intervals were randomly distributed) during a 150 s period. Only responses within 2 s following a probe were considered.

In the next part of the experiment, the secondary RT task was introduced. Each participant was asked to write three timed (7 min) paragraph-length narratives (describing their favorite book, movie, and TV show), in three different keyboard delay conditions (i.e., 0, 100, and 300 ms delay). The order of conditions and topics was counterbalanced. Participants were asked to react as quickly as possible to auditory probes during composition by pressing the button placed in front of the keyboard. Probes were presented on average every 30 seconds (15, 30, and 45 s randomly distributed intervals). Participants had a maximum of 2 s to react to each probe.

Results

A baseline RT for each participant was calculated as an average RT, with the first five trials (i.e., warm-up signals) excluded (Olive & Kellogg, 2002). The average RT in different typing conditions (i.e., 0, 100, and 300 ms delay) was calculated as an average RT with the first trial excluded (i.e., the first trial was always presented simultaneously with the task onset). Mean RTs across conditions are presented in Table 18. To address the violation of sphericity, the Greenhouse-Geisser correction was applied.

Table 18.

Baseline Reaction Time (RT), and 0, 100, and 300 ms RTs, Means and Standard Deviations (Experiment 5)

	Condition			
	Baseline	0 ms Delay	100 ms Delay	300 ms Delay
Mean (sec)	.62	1.01	1.02	1.04
Standard Deviation	.20	.17	.17	.19

A 4 level (baseline, and 0 ms, 100 ms, 300 ms keyboard delay) within subject ANOVA was used to assess the effect of condition on RT. There was a significant main effect of condition on RT, $F(1.85, 59.16) = 125.89$, $MSE = .02$, $p < 0.001$, $\eta^2_p = .80$. To compare the differences between RTs across conditions, a series of post-hoc tests was conducted using a Bonferroni correction. These analyses revealed that baseline RTs were reliably faster than RTs in the three typing conditions (i.e., 0, 100, and 300 ms delay), *Mean Differences* > .38, *ps* < .001, *ds* > 2.12. In contrast, RTs did not differ across typing conditions, *Mean Differences* < .04, *ps* > .65, *ds* < .18.

Discussion

The results of Experiment 5 clearly support the notion that slowing typing down by introducing the 100 ms (or 300 ms) keyboard delay did not markedly increase working memory demands during writing. Specifically, the dual-task effect was not larger in the keyboard delay conditions compared to standard typing. These findings provide strong support for the argument forwarded here to explain the discrepancy between the results reported in Chapters 1 and 3 and previous research. Namely, while previous research routinely attributed the relation between decreased transcription fluency and decreased lexical sophistication to the increased cognitive demands of disfluency, the results of Experiment 5 support the notion that some forms of disfluent transcription do not necessarily cause increased working memory demands. In such cases, increased disfluency could lead to increased lexical sophistication by allowing more time for lexical processes to unfold.

Conclusion

The current study provided a test between a time-based and a mapping disruption account of the effects of disfluency on written composition. The results were consistent with the former account. That is, by providing additional time for lexical processes to unfold, disfluency can increase lexical sophistication in written composition. However, the reported effects of disfluency on lexical sophistication did not translate to human-judged essay quality. Finally, I also demonstrated that the transcription fluency manipulation itself did not introduce large working memory demands.

In the next chapter I present a series of mediation analyses that provide a strong test of the time-based account, using a large sample of 420 essays (Chapters 1 and 3). Mediations were used to directly test an implied causal order, by which variation in transcription fluency across typing conditions should explain variation in lexical sophistication.

Chapter 4

In the following work I used a combined sample of 420 essays from studies described in Chapters 1 and 3 to test the implied causal sequence suggested by the time-based account, whereby variation in transcription fluency between typing conditions explains variation in lexical sophistication in essays.

According to the time-based account of the effects of disfluency on lexical sophistication, there is an implied causal order whereby variation in transcription fluency across typing conditions leads to variation in lexical sophistication. In other words, typing condition should influence lexical sophistication indirectly, through transcription fluency. I test this notion in the current chapter by performing a series of mediation analyses. In addition, another set of mediation analyses is performed to determine whether the effects of disfluency on lexical sophistication are related to differences in subjective fluency.

Transcription Fluency as the Mediator

I used a combined sample of 420 participants (Experiments 1, 3, and 4). Mediations were performed with typing condition (0 = standard, 1 = disfluent) as the predictor, transcription fluency as the mediator, and lexical sophistication indices (TTR or WF for content words, indices with the largest d in Medimorec & Risko, 2016, and Medimorec et al., 2017) as outcomes. I use the standard mediation annotation, where *a* = the unstandardized regression coefficient between condition and the mediator, *b* = the unstandardized regression coefficient between the mediator and the DV, controlling for the IV, *c* = the unstandardized regression coefficient between condition and the DV, and *c′* = the unstandardized regression coefficient between condition and the DV controlling for the mediator. Mediations were performed using the PROCESS modeling tool (Hayes, 2013), set to Model 4, with 1,000 bootstrap samples and a 95% confidence level for confidence intervals. The ratio of indirect to total effect (P_M; Hayes, 2013; Preacher & Hayes, 2008; Preacher & Kelley, 2011) is reported as the estimate of the effect size of the indirect effect.

Results

The results of the mediation analyses demonstrated that transcription fluency mediated the relation between typing condition and lexical sophistication indices. Table 19 illustrates the results of the mediation analyses.

Table 19.

Mediation Analyses: Transcription Fluency as the Mediator of the Relation Between Typing Condition (the Predictor) and Lexical Sophistication (i.e., Type-Token Ratio and Word Frequency for Content Words)

Outcome	Unstandardized regression coefficients				Indirect Effect [95% CI]	Effect Size
	a	b	c	c'		
Type-Token Ratio	107.20***	.0002**	.023***	.007	.016 [.006, .028]	.70
Word Frequency	107.20***	-.0003**	-.047***	-.011	-.036 [-.061, -.012]	.76

** $p < .05$ level. *** $p < .001$.

Subjective Fluency Measures as Mediators

A different account of the observed disfluency effects is that disfluency affects lexical sophistication indirectly, through subjective fluency. The results presented in Chapters 1 and 3 demonstrated that the transcription fluency manipulation also influenced the subjective fluency estimates (i.e., physical demand and frustration). Given previous evidence that subjective fluency can influence cognitive processing (e.g., Winkielman, et al., 2003), I used the combined sample to investigate the idea that the observed effects of disfluency on lexical sophistication could result from effects on participants' subjective fluency.

Mediations were performed using the same sample as above. As expected from the results of the individual experiments in Chapters 1 and 3, when the data were combined, participants in the disfluent conditions reported more mental and physical demand, as well as more effort and frustration, *Fs* > 4.50, *ps* < .035, *ds* > .20. The other measures (i.e., temporal demand and performance) were not affected, *Fs* < 1.31, *ps* > .24, *ds* < .12 (see Table 20).

Table 20.

Means, Standard Deviations, and Cohen's d of Subjective Fluency Measures for the Two Conditions

	Standard		Disfluent		Difference
Measure	M	SD	M	SD	Cohen's d
Mental Demand	-.28	4.59	.71	4.71	$.21^{**}$
Physical Demand	-6.51	4.04	-1.40	6.45	$.98^{***}$
Temporal Demand	-1.92	4.89	-1.60	5.32	.06
Performance	-2.84	4.63	-2.27	5.29	.11
Effort	.26	4.39	1.54	4.14	$.30^{**}$
Frustration	-3.87	5.24	-.55	6.10	$.59^{***}$

** $p < .05$ level. $^{***}p < .001$.

Thus, I performed a series of mediations with condition (0 = standard, 1 = disfluent) as the predictor, lexical sophistication indices (TTR or WF for content words) as the DV, and each of the affected TLX measures (mental, physical, demand, effort, and frustration) as an individual mediator. As is clear from Table 21, mediation models were not significant.

Table 21.

Relation Between the Condition and Lexical Sophistication Indices (Type-Token Ratio and Word Frequency for Content Words) Mediated by the Subjective Fluency Measures (Physical and Mental Demand, Effort, and Frustration)

		Unstandardized regression coefficients					
Outcome	Mediator	a	b	c	c'	Indirect Effect [95% CI]	Effect Size
Type-Token Ratio							
	Physical Demand	5.12^{***}	.0003	$.024^{***}$	$.022^{***}$	.001 [-.003, .006]	.06
	Frustration	3.32^{***}	.0005	$.024^{***}$	$.022^{***}$	.002 [-.001, .005]	.07
	Mental Demand	$.98^{*}$	-.0001	$.024^{***}$	$.024^{***}$	-.0001 [-.001, .001]	.004
	Effort	1.28^{*}	-.001	$.024^{***}$	$.025^{***}$	-.001 [-.003, .0004]	.03
Word Frequency Content Words							
	Physical Demand	5.12^{***}	.002	$-.053^{***}$	$-.062^{***}$	.009 [-.004, .022]	.17
	Frustration	3.32^{***}	-.0003	$-.053^{***}$	$-.052^{***}$	-.001 [-.010, .007]	.02
	Mental Demand	$.98^{**}$	-.002	$-.053^{***}$	$-.052^{***}$	-.001 [-.007, .001]	.03
	Effort	1.28^{**}	-.002	$-.056^{***}$	$-.053^{***}$	-.003 [-.009, .001]	.05

** $p < .05$ level. $^{***}p < .001$.

Discussion

The results of the mediation analyses demonstrated that transcription fluency mediated the relation between typing condition and the lexical sophistication measures. Thus, variation in lexical sophistication across typing conditions can be accounted for by variation in transcription fluency across typing conditions. Moreover, variation in lexical sophistication across typing conditions could not be accounted for by variation in subjective fluency measures across the conditions. These results are consistent with the time-based account of the effects of disfluency on lexical sophistication. In other words, slowing typing down provided additional time for lower frequency words and more different words to be activated. In the following chapter, I investigate whether effects of disfluency on lexical selection extend beyond composition, using a single word generation task.

Chapter 5

The results of Experiment 6 have been reported at the 46th Annual Ontario Psychology Undergraduate Thesis Conference (Edathodu, Z**.,** Medimorec, S., & Risko E. F., 2016) and in the undergraduate honors thesis by Zenusha Edathodu (University of Waterloo).

Some statistical test values differ slightly between this dissertation and the previously reported values because some minor errors in the raw data coding were addressed here.

The results of the studies described in Chapters 1 and 3 support the notion that decreasing transcription fluency in the context of essay writing leads to an increase in lexical sophistication. Specifically, I introduced transcription disfluency by having participants typewrite essays using one hand or a slightly delayed keyboard (vs. standard typing). Computational text analysis demonstrated that essays typed in disfluent conditions (i.e., one-handed typing and keyboard delay) were more lexically sophisticated. Namely, essays typed in disfluent conditions contained less frequent words and higher lexical diversity compared to essays typed in the standard condition. According to the proposed time-based account of the effects of disfluency on lexical sophistication, decreasing transcription fluency allowed more time for lexical processes (e.g., lexical access) to unfold, and individuals selected lower frequency words since it potentially benefits their writing. As discussed in Chapters 1 and 3, the time-based account is consistent with previous research that demonstrated increased time needed to access lower frequency words (Crowe, 1998; Fama et al., 1998; Forster & Davis, 1984). For example, Fama et al. (1998) reported that word frequency decreased as a function of time in a verbal fluency task. In other words, individuals first accessed a “ready store” of (more frequent) words, followed by infrequent words.

In Experiments 6 and 7 I sought to investigate the effects of disfluency in a word production task. In other words, are the effects of transcription disfluency limited to lexical selection in the context of written composition (i.e., producing an essay), or are they also present in a word generation task (i.e. production of words belonging to a category)? Answering these questions would provide more insight about the mechanism underlying the effects of disfluency in writing and the dynamics of lexical access in general.

According to the time-based account (proposed in Chapter 1) decreased transcription fluency provides additional time for lower frequency alternatives to be activated. While this time-based account provides a basis for lower frequency and more lexical items being activated, it does not explain why those words are also selected in composition. In the context of essay writing there is arguably utility in selecting lower frequency words. For example, rare words are considered to be more sophisticated (e.g., Crossley & McNamara, 2012), and typical essay scoring rubrics contain "word choice" as one of the criteria of good writing. Moreover, previous studies have consistently demonstrated a relation between increased lexical sophistication and higher writing grades (Crossley & McNamara, 2011, 2012; McNamara et. al, 2014). Therefore, essay writers could potentially benefit (i.e., improve their writing quality) by using more sophisticated language. Thus, according to the time-based account, transcription disfluency provides additional time for lower frequency alternatives to be activated, and in the context of composition individuals presumably strategically select lower frequency words since this could potentially benefit their writing. Clearly, these strategic effects should not be found when words are produced in isolation (i.e., in a single word generation task). For example, in a word fluency task, participants are asked to produce as many words as possible starting with a letter cue within a given time frame. Presumably, in such a task choosing the lower frequency alternatives does not have the same utility as in composition. This is because, at least from the participants' perspective, the total number of words produced represents the only measure of "success" or optimal performance in the task.

An alternative explanation of the effects of disfluency on lexical sophistication could be that decreased transcription fluency promotes selection of lower frequency words in different linguistic contexts (i.e., even outside of the composition context), for example by disrupting a habitual association between more fluent typing and rapid access to higher frequency words (this is a variant of the more extreme procedural account – disrupted access to finger-letter pairings - ruled out in Chapter 3). Thus according to this account, decreasing transcription fluency would provide more time for lower frequency alternatives and more alternatives to be activated, while also influencing (i.e., disrupting) the relation between fluent typing and selection of higher frequency words. Indeed, there is some evidence that typing fluency can influence lexical processing in the single word context. For example, Cerni, Velay, Alario, Vaugoyeau, and Longcamp (2016) reported that typing fluency influenced performance in a lexical decision task, such that expert typists took longer to identify visually presented difficult pseudowords (i.e., words with few bimanual transitions, given that cross-hand successive keystroke intervals are shorter than within-hand successive keystroke intervals; Rumelhart, & Norman, 1982) compared to easy pseudowords (i.e., those with many bimanual transitions), while non-experts showed the opposite pattern. Cerni et al. (2016) argued that typing (or the motor response in writing) is not isolated from orthographic processes and the repetitive performance of typing can interact with and influence lexical processes.

Thus while the time-based account proposed in Chapter 1 has a strategic component (in the context of composition) and thus does not predict disfluency effects in a word fluency task, the "habitual relation" account raises the interesting possibility that disfluency effects could be detected even outside of the essay writing context, for example in a written (i.e., typed) word generation task. Demonstrating an effect of disfluency in this task would support the notion that choosing lower frequency alternatives is related to slowing typing speed, regardless of the linguistic context (i.e., composition and single word generation). Such a result would also demonstrate that the strategic component of the time-based account proposed earlier requires rethinking.

Word Fluency Task

To investigate the effects of transcription disfluency in a relatively decontextualized word production task, I used a written word fluency test. Specifically, I used a variation of the phonemic fluency test. In phonemic fluency tests, participants are asked to produce as many words as possible from a category within a given time frame (e.g., Thurstone, 1938; Crowe, 1998). Typically, both the number of words generated and word frequency decline over time. For example, Crowe (1998) recorded verbal production for each 15 s time slice within the 60 s interval and reported a significant decrease in both measures as a function of time. Crowe (1998) suggested that individuals first access the most readily available (i.e., high frequency) words followed by lower frequency words. In the current study, I use a similar test to investigate whether the effects of disfluency extend to single word production by asking participants to produce as many words as possible starting with a letter cue within an 80 s timeframe.

Experiment 6

Method

Participants

Participants were 52 undergraduate students. Sample size was determined using the effect size of .40 (based on Medimorec & Risko, 2016), and power of .80 (Faul et al., 2009).

Design

A 2 (both-handed vs. one-handed typing condition) within subject design was used. The order of conditions was counterbalanced. Participants in the one-handed condition used their dominant or non-dominant hand when typing (the order was counterbalanced).

Stimuli and Apparatus

Participants, seated in front of a 24-in computer monitor, typed using a standard QWERTY keyboard. Participants were presented with 16 letter-cues (“T”, “A”, “O”, “I”, “N”, “S”, “H”, “D”, “L”, “C”, “M”, “W”, “F”, “P” and “B”) using the e-Prime software (Schneider, Eschman & Zuccolotto, 2002). The letters that are at the low end of letter frequency (e.g., “Q”, "Z", “X”) were not used. The order of presentation was randomized. The computer screen was recorded.

Procedure

Participants were asked to type as many words as possible beginning with a letter cue. There were 16 letter cues, each presented for 80 s. Eight of the trials were completed with standard typing, and another set of eight trials with one-handed typing. Following the standard procedure (e.g., Crowe, 1998), participants were instructed not to use proper nouns, numbers, repeated words, or use morphological variants of words (e.g., “tooth” and “teeth”). Participants were given a 3-minute practice session prior to each part of the task. I also collected subjective workload measures, not reported here.

Word Count and Word Frequency

Word counts and frequencies were calculated as an average for each 20 s interval of the word fluency task. The frequencies were extracted using the SUBTLEXus lexical database (Brysbaert & New, 2009).

Results

A series of 2 (Condition: standard vs. one-handed typing) x 4 (20 s intervals) repeated measures ANOVAs were performed. Misspelled words and words not found in the SUBTLEXus lexical database were removed before analyses. This resulted in the removal of 4.22% of words overall. When there was a violation of sphericity the Greenhouse-Geisser correction was applied. Partial eta squares and Cohen's d's are reported as measures of effect size where appropriate. There was no effect of condition order.

Word Count

Condition had a significant effect on the number of produced words, $F(1,51) = 5.25$, $MSE = .76$, $p = .026$, $\eta^2 = .09$, such that the number of words was higher in the standard compared to the one-handed condition. Moreover, the number of words significantly decreased across intervals, $F(1.92, 98.31) = 260.97$, $MSE = .76$, $p < .001$, $\eta^2 = .85$. There was an interaction between condition and time interval, $F(2.58,131.57) = 14.50$, $p < .001$, $\eta^2 = .22$. In a simple effect analysis, there was a significant effect of condition in the first (i.e., 20 s) time interval $t(51) = 4.89$, $p < .001$, $d = .53$, such that there were more words produced in the both hand condition compared to one hand condition. There were no differences across the other time intervals, all *ts* < .64, *ps* > .52, *ds* < .07 (see Table 22 for means and standard deviations).

Table 22.

Means, Standard Deviations, and Effect Sizes (Cohen's d) of the Number of Words for the Two Conditions (Experiment 6)

	Both Hands		One Hand		Effect Size
Time Interval (s)	M	SD	M	SD	d
1 (20)	5.80	1.46	5.05	1.37	.53***
2 (40)	3.85	1.12	3.81	.97	.03
3 (60)	3.27	.94	3.21	.91	.06
4 (80)	2.71	.94	2.76	.88	-.06

$^{***}p < .001$ level.

Word Frequency

Condition had a significant effect on word frequency, $F(1,51) = 5.53$, $MSE = .09$, $p = .023$, $\eta^2 = .10$, such that word frequency was lower in the one-handed condition than the both-handed condition. Moreover, there was an effect of time interval on word frequency such that it decreased across intervals, $F(3,153) = 26.35$, $MSE = 2.35$, $p < .001$, $\eta^2 = .34$. Finally, there was a significant interaction between condition and time interval, $F(3,153) = 2.73$, $MSE = .06$, $p = .046$, $\eta^2 = .05$ (see Table 23). In a simple effect analysis, there were no significant effects of condition on word frequency across the first 3 intervals (i.e., 20, 40, and 60-s), all *ts* < 1.67, *ps* > .10, *ds* < .24 while in the last interval (i.e., 80-s) word frequency was lower in the one-handed condition, $t(51) = 2.88$, $p = .006$, $d = .49$.

Next, I performed a series of simple effect analyses to investigate where the differences in word frequency occurred within each condition. In the both-handed condition, the first time interval was significantly higher than all of the other intervals, *ts* > 4.00, *ps* < .001, *ds* > .68. There were no differences among the other intervals, *ts* < .30, *ps* > .76, *ds* < .05. In the one-handed typing condition, the first time interval was significantly higher than all of the other intervals, *ts* > 6.24, *ps* < .001, *ds* > .72. Moreover, the second interval was higher compared to the last interval, $t(51) = 3.55$, $MSE = .05$, $p = .001$, $d = .48$, and the third interval was (marginally) higher than the last interval, $t(51) = 1.81$, $MSE = .06$, $p = .077$, $d = .29$. There was no difference between intervals 2 and 3, $t(51) = 1.44$, $MSE = .05$, $p = .156$, $d = .19$.

Table 23.

Means, Standard Deviations, and Effect Sizes (Cohen's d) of Word Frequency for the Two Conditions (Experiment 6)

	Both Hands		One Hand		Effect Size
Time Interval (s)	M	SD	M	SD	d
1 (20)	3.22	.29	3.23	.35	-.01
2 (40)	2.99	.40	2.97	.36	.05
3 (60)	2.98	.36	2.90	.36	.23
4 (80)	2.97	.36	2.80	.34	.49**

** $p < .05$ level.

Discussion

The results of Experiment 6 demonstrated that both the number of words and word frequency decreased as a function of time regardless of condition. Moreover, participants generated more words in the standard typing condition compared to the disfluent (i.e., one-handed typing) condition. Critically, the results of Experiment 6 demonstrated that participants produced less frequent words in the disfluent condition than in the standard typing condition. Such a result supports the notion that the selection of lower frequency words could potentially be explained by the disruption of the relation between more fluent typing and rapid access to higher frequency words caused by the disfluency manipulation. That is, typing more fluently might bias individuals to choose more frequent words, presumably a consequence of the habitual relation between fluent writing and rapid access to more frequent words. Introducing disfluency in writing could potentially disrupt this habitual relation. As discussed earlier, while this would be consistent with the time-based account, it also raises a question about the generalizability of the strategic component of that account. However, there was also an interaction between condition and time interval for word frequency, suggesting that the patterns of means are more complex than suggested by the "habitual relation" account. Specifically, in the both-handed typing condition word frequency sharply decreased at the second interval, and remained at a similar level afterwards, while in the one-handed condition word frequency linearly decreased across intervals.

To my knowledge, this is the first time that the effects of disfluency on word frequency have been reported outside of the essay-writing context. In Experiment 7 I sought to replicate the results of Experiment 6 and further investigate the nature of the reported interaction between condition and time interval for word frequency by switching to disfluency introduced by the keyboard delay (Chapter 3).

Experiment 7

Method

Participants

Participants were 60 undergraduate students. Sample size was determined using the effect size of .37 (based on the average standard deviation in the simple effect analysis for word frequency in Experiment 6), and power of .80.

Design

A 2 (standard vs. keyboard delay condition) within subject design was used. The order of conditions was counterbalanced.

Stimuli and Apparatus

The experiment was programmed in Python, and used the keyboard delay software. A slight modification compared to Experiment 6 was that the 16 letter cues were divided into two groups (i.e., lists) of 8 letter cues: list 1 ("N", "S", "M", "W", "D", "L", "H", "F"), and list 2 ("A", "O", "B", "E", "I", "T", "C", "P"). The order of lists was counterbalanced, and the order of individual letter cues within lists was randomized. The rest was the same as in Experiment 6.

Procedure

Same as in Experiment 6.

Word Count and Word Frequency

Same as in Experiment 6.

Results

Analysis followed that used in Experiment 6. Using the same criteria, I removed 6.35% of words. There was no effect of condition order.

Word Count

Condition had a significant effect on the number of produced words, $F(1,59) = 7.80$, $MSE = .82$, $p = .007$, $\eta^2 = .12$, such that the number of words was higher in the standard compared to the delayed condition. Moreover, the number of generated words significantly decreased across intervals, $F(1.89, 111.36) = 587.75$, $MSE = .66$, $p < .001$, $\eta^2 = .91$, and there was an interaction between condition and time interval, $F(2.60,153.24) = 8.53$, $p < .001$, $\eta^2 = .13$. In a simple effect analysis, there was a significant effect of condition in the first (i.e., 20-s) time interval $t(59) = 3.966$, $p < .001$, $d = .48$, such that there were more words produced in the standard compared to the keyboard delay condition. There were no differences across the other time intervals, all $ts < 1.65$, $ps > .10$, $ds < .20$ (see Table 24).

Table 24.

Means, Standard Deviations, and Effect Sizes (Cohen's d) of the Number of Words for the Two Conditions (Experiment 7)

	Standard		Delay		Effect Size
Time Interval (s)	M	SD	M	SD	d
1 (20)	6.01	1.47	5.37	1.21	.48***
2 (40)	3.53	1.00	3.50	.86	.03
3 (60)	2.96	.97	2.85	.79	.12
4 (80)	2.55	.82	2.40	.74	.19

***$p < .001$ level.

Word Frequency

There was no effect of condition on word frequency, $F(1,59) = 1.64$, $MSE = .14$, $p = .206$, $\eta^2 = .03$. There was an effect of time interval on word frequency such that it decreased across intervals, $F(3,177) = 32.09$, $MSE = .05$, $p < .001$, $\eta^2 = .35$. Finally, there was a marginally significant interaction between condition and time interval, $F(3,177) = 2.21$, $MSE = .08$, $p = .089$, $\eta^2 = .04$ (for means and standard deviations see Table 24). In a simple effect analysis, word frequency was lower in the standard condition in the second interval (i.e., 40 s), $t(59) = 2.61$, $p = .011$, $d = .41$. There were no differences across the other intervals, all *ts* < 1.21, *ps* > .23, *ds* < .21 (see Table 25 and Figure 1).

Next I performed a series of simple effect analyses within each condition. In the standard condition, the first time interval was significantly higher than all of the other intervals, *ts* > 4.37, *ps* < .001, *ds* > .56. There were no differences among the other intervals, (absolute) *ts* < .84, *ps* > .40, *ds* < .13. In the delayed keyboard condition, the first time interval was significantly higher than all of the other intervals, *ts* > 4.39, *ps* < .001, *ds* > .60. Moreover, the second interval was (marginally) higher compared to the third interval, $t(59) = 1.87$, $MSE = .04$, $p = .066$, $d = .22$, and higher than the fourth interval, $t(59) = 2.64$, $MSE = .05$, $p = .011$, $d = .45$. There was no difference between intervals 3 and 4, $t(59) = 1.37$, $MSE = .05$, $p = .177$, $d = .21$.

Table 25.

Means, Standard Deviations, and Effect Sizes (Cohen's d) of Word Frequency for the Two Conditions (Experiment 7)

	Standard		Delay		Effect Size
Time Interval (s)	M	SD	M	SD	d
1 (20)	3.23	.37	3.29	.28	-.20
2 (40)	2.98	.32	3.11	.31	-.41**
3 (60)	3.01	.38	3.04	.34	-.07
4 (80)	3.02	.33	2.97	.31	.15

$^{**}p < .05$ level.

Discussion

The results of Experiment 7 demonstrate that the number of generated words and word frequency decreased as a function of time in both conditions. Again, participants generated more words in the standard typing condition than in the disfluent condition. The results of Experiment 7 were different from those reported in Experiment 6 in that here transcription fluency had no overall statistically significant effect on word frequency (i.e., there were no differences in word frequency between the conditions). However, there was a marginal interaction between condition and time interval for word frequency. This interaction was also observed in Experiment 6.

Indeed, a closer examination of the word frequency means as a function of time interval indicates similar functions between corresponding conditions across the two experiments (the only difference being a relative upward shift in the disfluent condition here compared to Experiment 6; see Figure 1). In the standard condition of both experiments word frequency quickly drops at the second (i.e., 40 s) interval, but remains relatively flat afterwards. The simple effects analysis demonstrated that word frequency was higher in the first interval compared to the rest, and there were no differences among the other intervals (i.e., 20, 40, and 80 s; note that this replicates previously reported results; e.g., Crowe, 1998). In the disfluent conditions, in contrast, word frequency linearly decreased across intervals. The simple effect analysis demonstrated that the first time interval was significantly higher than all of the other intervals. In addition, there were differences among the other intervals as well (i.e., the second and third (marginally) intervals were higher compared to the last interval in Experiment 6, while the second interval was higher compared to the third (marginally) and fourth intervals). Thus the observed patterns are clearly more complex than either of the proposed accounts would have suggested. I examine the implications of these results further in the General Discussion.

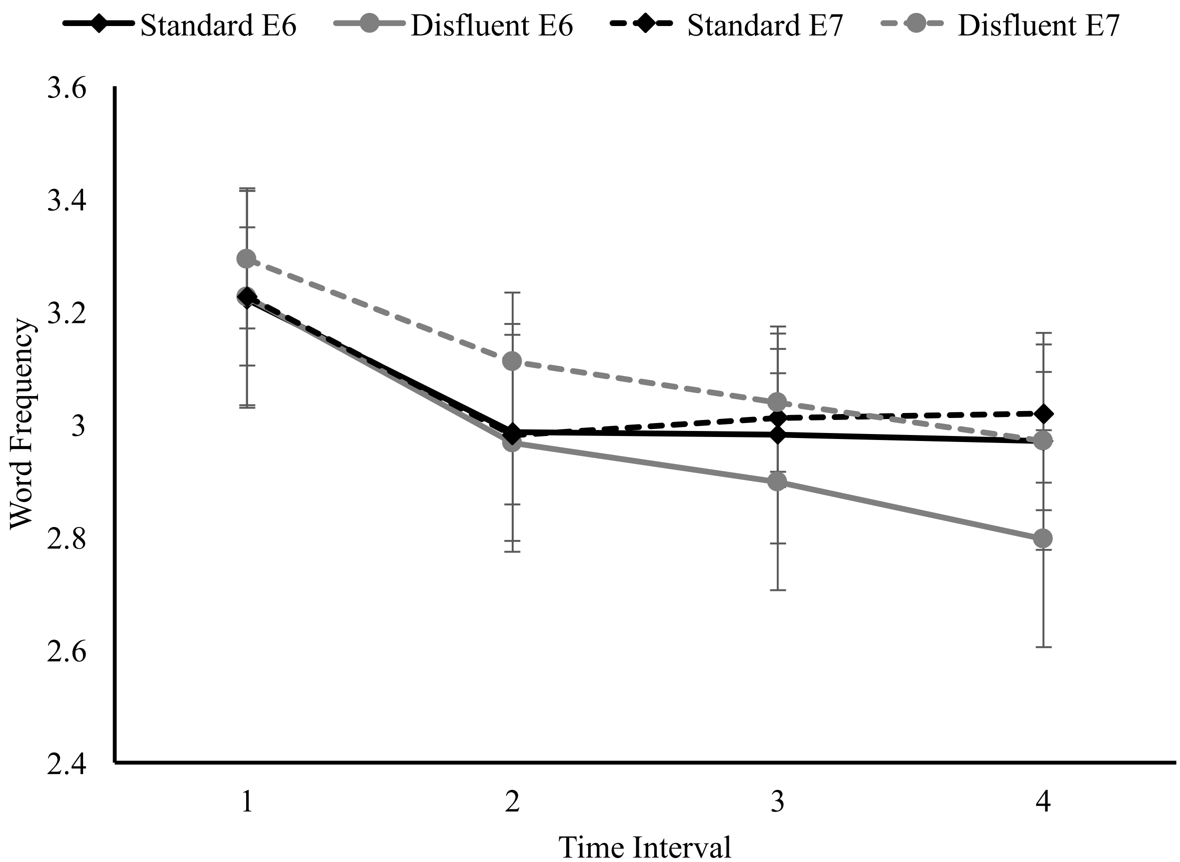

Figure 1. Word frequency as a function of time for Standard and Disfluent typing conditions in Experiment 6 (E6) and Experiment 7 (E7). The error bars represent 95% Loftus and Masson (1994) confidence intervals.

General Discussion

Both experiments replicated previous studies that reported a decrease in the number of generated words and word frequency as a function of time (e.g., Crowe, 1998; Fama et al., 1998). This was true regardless of condition (i.e., standard vs. disfluent typing). However, there were also some differences between the findings of the two experiments. For example, in Experiment 6 participants overall produced lower frequency words in the disfluent typing condition (i.e., one-handed typing) compared to the standard typing condition. This effect of disfluency on word frequency was not replicated in Experiment 7. However, visual inspection of the function relating word frequency and time interval (see Figure 1) reveals, on the other hand, a consistent pattern wherein in the standard condition, across the two experiments, word frequency steeply declines at the second interval, but remains relatively flat afterwards. On the other hand, in the disfluent conditions word frequency linearly decreases across time intervals. The results of the simple effect analyses confirmed these different patterns. The pattern in the standard condition is consistent with previous research (Crowe, 1998). Thus, it appears as though the disfluency manipulations influence lexical selection in the word fluency task but not in as straightforward a manner as suggested by the proposed accounts (i.e., either no effect or a main effect). I speculate below on why this pattern might have emerged.

Performance in word fluency tasks is typically taken to reflect the initial production of readily available (i.e., automatically activated) frequently used words in the first 15-20 s of the task, followed by a more effortful and less productive search during the subsequent intervals (Crowe, 1998; Hurks et al., 2005; Unsworth, 2017). For example, recent research has demonstrated that individuals use various strategies across word fluency tasks (Unsworth, 2017). Unsworth asked participants to produce as many words as possible from a category (e.g., animals, super market items), and in addition introduced thought-probes asking participants what strategy they were using to retrieve items. Participants most frequently reported using visualization (i.e., imagining various activities, locations), link-to-previous (i.e., using the last generated item to think of new items), and relying on no strategy. Interestingly, participants also frequently reported using no strategy at the first probe, followed by an adoption of a more active strategy later on in the task. Thus it seems that participants often start off with no particular strategy, which supports the notion that the more frequent words are spontaneously and passively retrieved (Hills et al., 2013). The observed effects of disfluency in the current study could reflect the disfluency manipulation influencing strategy choice in the word fluency task. For example, staying within the same frequency band starting with the second interval could indicate that individuals settled on one strategy, while decreasing word frequency could indicate a more dynamic change in search strategies. Of course, the explanation forwarded here is speculative and future research investigating the effects of disfluency in a single word production task is necessary.

Future studies investigating this question could use a larger sample, and provide more time per letter cue (compared to 80 s here), thus allowing participants to generate more words. Based on the results of the current study, across additional time intervals we should expect a relatively flat function in the standard condition, and a downward trend in the disfluent condition. Moreover, introducing thought-probes asking participants about their strategy during the task could provide more information about the mechanisms underlying disfluency effects in this task.

What do the present results suggest with respect to interpreting the effects of disfluency in composition reported earlier? Unfortunately, the results leave us in a somewhat ambiguous position because neither prediction was clearly supported. While overall the effect of disfluency on word frequency was clearly less robust than in composition as predicted by the time based account I outlined (wherein participants select lower frequency words if they are available), it is hard to argue that there was no effect of disfluency on lexical selection. Namely, there was a main effect in Experiment 6 and across experiments there was a clear difference in the functions relating time and word frequency across the fluent and disfluent conditions. If future work clears up the empirical pattern, then stronger conclusions might be able to be drawn. Finally, it is worth noting the possibility that the processes underlying lexical selection in composition and a single word fluency task are not equivalent and as such any disfluency effect observed in these tasks may not be the product of the same influence.

Chapter 6

Concluding Remarks

In this series of studies, I have demonstrated that decreased transcription fluency (or typing speed) can lead to increased lexical sophistication in the context of essay writing. In Chapter 1, I introduced disfluency by asking participants to type essays using one hand (vs. standard typing). The results demonstrated that decreasing transcription fluency resulted in more lexically sophisticated essays. Specifically, essays written in the disfluent condition contained less frequent words and were more lexically diverse compared to standard typing. The reported results are surprising because transcription fluency is usually thought to be positively related to writing quality as the automatization of transcription arguably affords the re-distribution of resources to higher level writing processes (e.g., planning; Fayol, 1999; Kellogg, 1999). To explain the observed effects of disfluency in composition in Chapter 1, I proposed the time-based account of disfluency in composition whereby decreasing transcription fluency allows more time for lexical processes, such as lexical access, to unfold, followed by the strategic choice of more lexically sophisticated words. It is also important to note that the effects of disfluency described in Chapter 1 were observed when participants typed longer (approximately 600 word) essays (narrative or argumentative). However, the effect of disfluency was not observed when participants composed shorter (approximately 300 word) essays, presumably because with the shorter essays individuals did not have sufficient practice to familiarize themselves with disfluent (i.e., one-handed) typing. Such an explanation is supported by subjective fluency measures, since participants who wrote shorter essays reported more overall effort compared to participants who wrote longer essays.

I continued the investigation of the effects of disfluency in composition in Chapter 2, by investigating the extent to which introducing a disfluency manipulation influences pauses and revisions in composition. The results demonstrated that less fluent typing is not related to increased pause rates. I also demonstrated no relation between the experimental manipulation of transcription fluency and the number of revisions in composition.

In addition to decreasing transcription fluency, typing with one hand also disrupts habitual finger-to-letter mappings, which could in theory influence lexical selection during writing (Pelleg et al., 2015). In Chapter 3, I provided a strong test between the two accounts of the effects of disfluency in writing, the time-based account, which attributes the effects to the delay caused by the disfluency, and an account that attributes the effects to the disruption of typical finger-to-letter mappings caused by the disfluency. To test between these accounts, I slowed down participants' typing by introducing a small delay (i.e., 100 ms) between consecutive keystrokes while individuals typed essays. Critically, this manipulation did not disrupt typical finger-to-letter mappings. The study presented in Chapter 3 replicated the results from Chapter 1. Namely, the essays written in the disfluent condition were again more lexically diverse and contained less frequent words. These results are consistent with the time-based account (i.e., decreasing transcription fluency allows more time for lexical processes to unfold). As in Chapter 1, these results also diverge from previous research that typically reported a positive relation between transcription fluency and lexical sophistication (e.g., Olive et al., 2009). In Chapter 3 I also tested the hypothesis that, unlike in previous studies, the transcription disfluency manipulation in the current study did not introduce large working memory demands. To investigate this notion, I performed another experiment, where I used a dual-task technique to investigate working memory demands of the introduced disfluency. The results of this experiment demonstrated no difference between standard and disfluent typing conditions with respect to participants' response time to pseudorandom probes presented during composition.

Thus, the results of the experiment supported the notion that the transcription disfluency manipulation (i.e., keyboard delay) did not introduce large working memory demands. Finally, the results presented in Chapter 3 also demonstrated that the reported differences in lexical sophistication did not translate to an effect on human assessed essay quality. I speculated that there could be several potential reasons for this result, including a relatively low reliability between essay graders, and a magnitude of the effect of transcription disfluency on lexical sophistication that is insufficient to translate into grader-assessed essay quality. Future work is required to better understand why the increase in lexical sophistication due to the disfluency does not translate to greater human ratings of quality (given the well-known relation between the two variables).

The time-based account of the effects of disfluency in composition was further supported by the results of mediation analyses presented in Chapter 4. I used mediations to test the implied causal order whereby typing condition should influence lexical sophistication (i.e., word frequency and lexical diversity) indirectly, through transcription fluency, measured by the mean keystroke interval within a word. Mediation analyses were performed using a large sample of 420 essays from Chapters 1 and 3. The results of the analyses indeed demonstrated that transcription fluency mediated the relation between typing condition and lexical sophistication. In other words, mediations strongly supported the notion that variation in lexical sophistication across typing conditions can be accounted for by variation in transcription fluency. These results are consistent with the time-based account. Furthermore, since participants in disfluent conditions reported more mental and physical demand, as well as more effort and frustration, I performed another set of mediations to investigate whether the relation between typing condition and lexical sophistication is mediated by differences in subjective fluency. The results demonstrated that variation in lexical sophistication across typing conditions could not be accounted for by variations in subjective fluency measures.

I further extended the investigation of transcription disfluency in Chapter 5, where I sought to investigate whether effects of disfluency on lexical selection extend beyond composition to a single word generation task. In this task, I asked participants to produce as many words as possible starting with a letter cue within an 80 s interval. The results of these experiments failed to provide clear evidence for any of the proposed accounts. To explain these results, I speculated that introducing disfluency might have influenced a strategy used to retrieve words. That said, more work is needed using the word fluency task together with a disfluency manipulation. In addition, it would be prudent to examine the relation between lexical selection in composition and lexical selection in word fluency tasks (e.g., is there a correlation between the two?).

Implications for our Understanding of Lexical Selection in Composition

Much previous research investigating lexical selection and the development of various models of lexical selection is based on relatively decontextualized naming paradigms (e.g., variations of a naming task; Caramazza & Costa, 2000; 2001; Dhooge & Hartsuiker, 2011; Finkbeiner & Caramazza, 2006; Mahon, Costa, Peterson, Vargas & Caramazza, 2007; Morsella & Miozzo, 2002; Navarrete & Costa, 2005; Roelofs, 1992), without much consideration for peripheral processes such as the dynamics of transcription fluency or speech rate and their potential influence on lexical choice (e.g., Strijkers, & Costa, 2016). In the current work I have investigated lexical selection in a relatively unconstrained essay writing task, and a more constrained word fluency task, while also manipulating transcription fluency. Critically, the results of the current work indicate that word production can be influenced by interfering with transcription fluency.

I interpreted the influence of an introduced disfluency on lexical sophistication by suggesting that the disfluency allowed additional time for the activation of a larger number of words, and lower frequency words. This time based activation of additional lexical items could be realized through time dependent spreading activation mechanisms (Collins & Loftus, 1975; Levelt, Roelofs, & Meyer, 1999; Roelofs, 1992). For example, much previous research has demonstrated that frequent words are accessed faster than infrequent words (e.g., Forster & Chambers, 1973; Knobel et al., 2008). But how do these additional items (relative to my fluent transcription condition) ultimately get selected during composition? Most theories of lexical selection posit that the strongest candidate for selection is either the one with the highest activation (i.e., competitive selection; e.g., Howard, Nickels, Coltheart, & Cole-Virtue, 2006) or simply the one that first exceeds a pre-determined absolute activation threshold (i.e., a non-competitive selection process; e.g. Mahon, Costa, Peterson, Vargas, & Caramazza, 2007).

One potential explanation for the effects of disfluency is to suggest that individuals in the disfluent condition might be better able to override an initially selected response (e.g., the most active item). For example, Nozari, Freund, Breining, Rapp, and Gordon (2016) argued that a controlled process in spoken word production can suppress a potent but less preferred response before its overt production. Nozari et al. (2016) referred to this process as post-monitoring control. Thus one potential way to explain word selection in the disfluent conditions is to suggest that introducing disfluency allowed additional time for this post-monitoring control to operate. In such a case individuals might be better able to suppress words (e.g., high frequency words) that are more highly activated or that cross some threshold first, thus allowing a lexical search to continue, for example, until a better fit to the context reaches the highest activation level or a more "sophisticated" word is discovered. From this perspective, lexical selection in composition reflects a competition between the relatively passive activation of lexical options and the more strategic processes related to post-monitoring control (e.g., wanting to select "sophisticated" words, or the "best-fitting" word). Introducing disfluency might aid the latter processes in that competition. Consistent with this notion, there is evidence that individuals slow down speech rate when producing lower frequency words (e.g., Cohen Priva, 2017). On this account the importance of the disfluency not burdening working memory is also clear since post-monitoring control is likely resource demanding. As such, a disfluency manipulation that increased WM demands would likely not increase lexical sophistication despite the fact that it also provides more time for lexical access.

The results of the current study also raise an interesting question about typical typing. Specifically, could individuals simply be instructed to slow down their typing (e.g., in standard typing conditions in this work) to increase their lexical sophistication? It seems unlikely that such an instruction would be effective, because skilled typing represents a routine (or automatic) action (e.g., Cerni, Longcamp, & Job, 2016; Crump & Logan, 2010) and inhibition of relatively automatic actions is costly since it requires attentional resources. For example, previous research has demonstrated that paying attention to the mechanics of typing had disruptive effects, such that individuals, for example, made more errors (Logan & Crump, 2009). Thus, intentionally slowing typing would require attentional resources and could arguably be detrimental to writing.

Implications for Writing Implements

Do writing implements influence our writing? Individuals have often intuited so, usually attributing potential changes in writing to a variation in transcription fluency across writing tools.

For example, Norman (2002) has suggested that increasing writing speed by switching from handwriting to typewriting should lead to writing with “less thought and care” (p. 210). Presumably, even more fluent writing should start to resemble everyday speech (Norman, 2002). Consistent with this idea, in the current series of studies I provided evidence that more fluent transcription in typing is related to decreased lexical sophistication of essays. Thus the results of the current study underline the importance of thinking about writing as a kind of extended cognitive system (Menary, 2007, 2010). On this view writing is a product of the interaction between both internal processes and external or peripheral ones (i.e., writing implements). Thus, changes in parts of the system that ostensibly should not influence what is written (e.g., the output modality) can and do through influencing the operation of the internal processes governing, for example, word selection. In other words, various writing tools affording different writing speeds presumably interact with the human mind in different, but predictable ways. This is interesting to consider in light of moves toward speech-to-text systems. For example, consistent with the results presented in this work, there is recent evidence that a fast speech rate is correlated with the use of less informative (i.e., more frequent) words, at least in conversational contexts (Cohen Priva, 2017).

The above question about writing implements, as well as the idea of writing in general, is intimately related to the concept of technology. Indeed, Ong (2002) has argued that writing itself is “a technology, calling for the use of tools” (p. 80). Interestingly, the technology of writing has been evolving through gradual automation of writing tools (and thus automation of transcription; Norman, 2002). One of the profound changes in writing technology includes the invention of the typewriter. At the time, the Scientific American editorial somewhat prophetically proclaimed that “The weary process of learning penmanship in schools will be reduced to the acquirement of the writing of one's own signature." (as quoted in Wershler-Henry, p. 143). Almost immediately, individuals started intuiting that switching from pen and paper to the more automatic typewriting brought about changes in their writing. My dissertation begins with the famous Nietzsche quote echoing such a sentiment. Clearly, as the technology of writing keeps on changing, it is important to understand the impact of those changes on writing processes. The current series of studies has provided initial evidence that automation of transcription can have detrimental effects on writing processes. Specifically, more fluent typing was related to decreased lexical sophistication in the context of essay writing. Interestingly, one of the consequences of the gradual automation of the writing tools is that at this stage of technological development transcription speed is only limited by the rate of speech (i.e., speech-to-text composition). Such a high transcription fluency could represent what Norman (2002) referred to as *overautomation*, or "too great a degree of automation" (p. 195). Perhaps ironically, producing written language by using the most automated technology available today, at least on the face of it resembles a less sophisticated oral composition that predates literacy.

Future Directions

In this dissertation I have demonstrated the existence of rich interactions between internal writing processes and how individuals write. Clearly, introducing transcription disfluency influenced individuals' word choice during essay writing. Thus, the evidence presented here supports the notion that disfluency can lead to increased lexical sophistication. A number of theoretical predictions derived from the current study offer different avenues that future studies could explore. One of those questions is whether disfluency effects can be detected by comparing composition across different writing tools given differences in composition speed across the tools (e.g., speech-to-text should presumably be more fluent compared to typing, while typing is more fluent compared to longhand writing). Thus future studies about the effects of transcription disfluency could provide additional insights into mechanisms underlying language production.

References

Alves, R.A., Castro, S.L., de Sousa, L., & Strömqvist, S. (2007). Influence of typing skill on pause–execution cycles in written composition. In Rijlaarsdam, G. (Series Ed.); M. Torrance, L. van Waes, & D. Galbraith (Volume Eds.), *Writing and Cognition: Research and Applications* (Studies in Writing, Vol. 20, pp. 55–65). Amsterdam, The Netherlands: Elsevier.

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*, 969-979.

Avent, J. & Austermann, S. (2003). Reciprocal scaffolding: A context for communication treatment in aphasia. *Aphasiology, 17,* 397–404.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1996). Celex2. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Ball, B. H., Klein, K. N., & Brewer, G. A. (2014). Processing fluency mediates the influence of perceptual information on monitoring learning of educationally relevant materials. *Journal of Experimental Psychology: Applied, 20,* 336–348.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum.

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89*, 1-47.

Breland, H., Lee, Y. W., & Muraki, E. (2005). Comparability of TOEFL CBT essay prompts: response-mode analyses*. Educational and Psychological Measurement, 65*, 577-595.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language, 44,* 274–296.

Bridgeman, B., & Cooper, P. (1998, April). *Comparability of scores on word-processed and handwritten essays on the graduate management admissions test.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

British National Corpus Consortium. (2007). British national corpus version 3 (BNC XMLth ed.).

Brown, C. M. (1988). Comparison of typing and handwriting in “two-finger typists.” *Proceedings of the Human Factors Society, 32*, 381–385.

Butterfield, E. C., Hacker, D. J., & Plumb, C. (1994). Topic knowledge, linguistic knowledge, and revision processes as determinants of text revision. *Advances in Cognition and Educational Practice*, *2*, 83-141.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977-990.

Butler, J. A., & Britt, M. A. (2011). Investigating instruction for improving revision of argumentative essays. *Written Communication, 28*, 70-96.

Cain, K., & Nash, H. M. (2011). The influence of connectives on young readers’ processing and comprehension of text. *Journal of Educational Psychology, 103,* 429–441.

Caramazza, A., & Costa, A. (2000). The semantic interference effect in the picture-word interference paradigm: does the response set matter? *Cognition*, *75*, B51-B64.

Caramazza, A., & Costa, A. (2001). Set size and repetition in the picture–word interference paradigm: Implications for models of naming. *Cognition*, *80*, 291-298.

Cerni, T., Longcamp, M., & Job, R. (2016). Two thumbs and one index: A comparison of manual coordination in touch-typing and mobile-typing. *Acta Psychologica*, *167*, 16-23.

Cerni, T., Velay, J. L., Alario, F. X., Vaugoyeau, M., & Longcamp, M. (2016). Motor expertise for typing impacts lexical decision performance. *Trends in Neuroscience and Education*, *5*, 130-138.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.) Hillsdale, NJ: Erlbaum.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407-428.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33*, 497–505.

Connelly, V., Campbell, S., MacLean, M., & Barnes, J. (2006). Contribution of lower order skills to the written composition of college students with and without dyslexia. *Developmental Neuropsychology*, *29*, 175-196.

Connelly, V., Dockrell, J. E., & Barnett, J. (2005). The slow handwriting of undergraduate students constrains overall performance in exam essays. *Educational Psychology*, *25*, 99-107.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2015). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods,* 1-11.

Crossley, S. A., & McNamara, D. S. (2011). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life Long Learning, 21,* 170-191.

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The role of cohesion, readability, and lexical difficulty. *Journal of Research in Reading, 35,* 115-135.

Crossley, S., Salsbury, T., & McNamara, D. (2010). The development of polysemy and frequency use in English second language speakers. *Language Learning*, *60*, 573-605.

Crossley, S. A., Weston, J., McLain Sullivan, S. T., & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis *Written Communication, 28*, 282-311.

Crowe, S. F. (1998). Decrease in performance on the verbal fluency test as a function of time: Evaluation in a young healthy sample. *Journal of Clinical and Experimental Neuropsychology*, *20*, 391-401.

Crump, M. J., & Logan, G. D. (2010). Hierarchical control and skilled typing: evidence for word-level control over the execution of individual keystrokes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1369-1380.

Deane, P., & Quinlan, T. (2010). What automated analyses of corpora can tell us about students' writing skills. *Journal of Writing Research, 2,* 151-177.

Dhooge, E., & Hartsuiker, R. J. (2011). The distractor frequency effect in a delayed picture-word interference task: Further evidence for a late locus of distractor exclusion. *Psychonomic Bulletin & Review, 18*, 116-122.

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition, 118*, 111–115.

Duran, N.D., McCarthy, P.M., Graesser, A.C. & McNamara, D.S. (2007). Using temporal cohesion to predict temporal coherence in narrative and expository texts. *Behavior Research Methods*, *29*, 212–223.

Edathodu, Z., Medimorec, S., & Risko E. F. (2016, May). *The effect of disfluency on electronic word fluency test.* Presented at the 46th Annual Ontario Psychology Undergraduate Thesis Conference, Waterloo, ON, Canada.

Fama, R., Sullivan, E. V., Shear, P. K., Cahn-Weiner, D. A., Yesavage, J. A., Tinklenberg, J. R., & Pfefferbaum, A. (1998). Fluency performance patterns in Alzheimer's disease and Parkinson's disease. *The Clinical Neuropsychologist*, *12*, 487-499.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160.

Fayol, M. (1999). From on-line management problems to strategies in written composition. In M. Torrance & G. Jeffery (Eds.), *The cognitive demands of writing: Processing capacity and working memory effects in text production* (pp. 13–23). Amsterdam, The Netherlands: Amsterdam University Press.

Finkbeiner, M., & Caramazza, A. (2006). Now you see it, now you don't: On turning semantic interference into facilitation in a Stroop-like task. *Cortex*, *42*, 790-796.

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

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*, 627-635.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 680-698.

Gentner, D. R. (1988). Expertise in typewriting. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 1–21). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Gilhooly, K.J. & Logie, R.H. (1980). Age of acquisition, imagery, concreteness, familiarity and ambiguity measures for 1944 words. *Behaviour Research Methods and Instrumentation*, *12,* 395–427.

Gonzalez, M.C. (2014, November). *Lexical diversity, sophistication, and size in academic writing*. Presented at the 13th Annual Symposium on Second Language Writing, Tempe, AZ.

Grabowski, J. (2007). The writing superiority effect in verbal recall of knowledge: Sources and determinants. In M. Torrance, L. van Waes, & D. Galbraith (Eds.), *Writing and cognition: Research and applications* (pp. 165-179). Amsterdam, The Netherlands: Elsevier.

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science, 3,* 371–398.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher, 40,* 223–234.

Graesser, A. C., McNamara, D. S., & Louwerse, M. M. (2003). What do readers need to learn in order to process coherence relations in narrative and expository text? In A. P. Sweet & C. Snow (Eds.), *Rethinking reading comprehension* (pp. 82–98). New York, NY: Guilford.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments & Computers, 36,* 193–202.

Graham, S., Berninger, V. W., Abbott, R. D., Abbott, S. P., & Whitaker, D. (1997). Role of mechanics in composing of elementary school students: A new methodological approach. *Journal of Educational Psychology*, *89*, 170-182.

Graham, S., Gillespie, A., & McKeown, D. (2013). Writing: Importance, development, and instruction. *Reading and Writing*, *26*, 1-15.

Grela, B.G. (2002). Lexical diversity in children with Down Syndrome. *Clinical Linguistics and Phonetics*, *16*, 251–263.

Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing*, *18*, 218-238.

Heidegger, M. (1992). *Parmenides* (A. Schuwer & R. Rojcewicz, Trans.). Bloomington, IN: Indiana University Press.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139-183.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* New York, NY. Guilford Press.

Hills, T. T., Mata, R., Wilke, A., & Samanez-Larkin, G. R. (2013). Mechanisms of age-related decline in memory search across the adult life span. *Developmental Psychology, 49*, 2396.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, *100*, 464-482.

Hurks, P. P. M., Hendriksen, J. G. M., Vles, J. S. H., Kalff, A. C., Feron, F. J. M., Kroes, M., ... & Jolles, J. (2004). Verbal fluency over time as a measure of automatic and controlled processing in children with ADHD. *Brain and Cognition*, *55*, 535-544.

Inhoff, A. W. (1991). Word frequency during copytyping. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 478-487.

Karat, C. M., Halverson, C., Horn, D., & Karat, J. (1999, May). Patterns of entry and correction in large vocabulary continuous speech recognition systems. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 568-575). ACM.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773-795.

Kellogg, R. T. (1994). *The psychology of writing*. New York, NY: 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 application* (pp. 57–71). Mahwah, NJ: Lawrence Erlbaum Associate Inc.

Kellogg, R. T. (1999). Components of working memory in text production. In M. Torrance & G. Jeffery (Eds.), *The cognitive demands of writing: Processing capacity and working memory effects in text production* (pp. 43–61). Amsterdam, The Netherlands: Amsterdam University Press.

Kellogg, R. T. (2001). Competition for working memory among writing processes. *The American Journal of Psychology*, *114*, 175-191.

Kellogg, R. T., Olive, T., & Piolat, A. (2007). Verbal, visual, and spatial working memory in written language production. *Acta Psychologica*, *124*, 382-397.

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

Klein, W. (1994). *Time in language.* London, UK: Routledge.

Knobel, M., Finkbeiner, M., & Caramazza, A. (2008). The many places of frequency: Evidence for a novel locus of the lexical frequency effect in word production. *Cognitive Neuropsychology*, *25*, 256-286.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly, 49*, 757-786.

Larsen-Freeman, D. (2002). Making sense of frequency. *Studies in Second Language Acquisition, 24,* 275–285.

Laufer, B. (1994). The lexical profile of second language writing: Does it change over time? *RELC journal, 25*, 21-33.

Laufer, B. & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16,* 307-322.

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

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-38.

Levy, C. M., & Ransdell, S. E. (1995). Is writing as difficult as it seems? *Memory and Cognition, 23*, 767–779.

Lindqvist, C., Gudmundson, A., & Bardel, C. (2013). A new approach to measuring lexical sophistication in L2 oral production. In Bardel C., Lindqvist, C. & Laufer B., (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 109–126). Eurosla Monographs Series 2. Retrieved from http://eurosla.org/.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476–490.

Logan, G. D., & Crump, M. J. (2009). The left hand doesn't know what the right hand is doing: The disruptive effects of attention to the hands in skilled typewriting. *Psychological Science, 20*, 1296-1300.

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., Ly, A., Gronau, Q. F., Smira, M., Epskamp, S., Matzke, D., Wild, A., Rouder, J. N., Morey, R. D. & Wagenmakers, E.-J. (2015). JASP (Version 0.7) [Computer software].

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 503-535.

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

Matias, E., MacKenzie, I. S., & Buxton, W. (1996). One-handed touch-typing on a QWERTY keyboard. *Human-Computer Interaction, 11*, 1-27.

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,* 381-392.

McCutchen, D. (1988). Functional automaticity in children's writing: A problem of metacognitive control. *Written Communication, 5,* 306-324.

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review*, *8*, 299-325.

McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational Psychologist*, *35*, 13-23.

McCutchen, D., Covill, A., Hoyne, S. H., & Mildes, K. (1994). Individual differences in writing: Implications of translating fluency. *Journal of Educational Psychology*, *86*, 256-266.

McNamara, D. S. (2013). The epistemic stance between the author and reader: A driving force in the cohesion of text and writing. *Discourse Studies*, *15*, 579-595.

McNamara, D.S., Crossley, S.A., & McCarthy, P.M. (2010). Linguistic features of writing quality. *Written Communication*, *27*, 57–86.

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

Medimorec, S., Pavlik, P. I., Jr., Olney, A., Graesser, A. C., & Risko, E. F. (2015). The language of instruction: Compensating for challenge in lectures. *Journal of Educational Psychology, 107,* 941-990.

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

Medimorec, S., & Risko, E. F. (2017). Pauses in written composition: On the importance of where we pause. *Reading and Writing.* doi:10.1007/s11145-017-9723-7

Medimorec, S., Young, T. P., & Risko, E. F. (2017). Disfluency effects on lexical selection. *Cognition, 158,* 28-32.

Menary, R. (2007). Writing as thinking. *Language Sciences, 29,* 621–632.

Menary, R. (2010). Dimensions of mind. *Phenomenology and the Cognitive Sciences, 9*, 561–578.

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 555-563.

Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological Science*, *25*, 1-10.

Nadler, R. T., Rabi, R., & Minda, J. P. (2010). Better mood and better performance learning rule-described categories is enhanced by positive mood. *Psychological Science, 21,* 1770–1776.

Navarrete, E., Basagni, B., Alario, F. X., & Costa, A. (2006). Does word frequency affect lexical selection in speech production? *The Quarterly Journal of Experimental Psychology, 59,* 1681-1690.

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language*, *53*, 359-377.

Neuwirth, C. M., Haas, C., & Hayes, J. R. (1990) *Does word processing improve students' writing? A critical appraisal and assessment* (Final Rep. to FIPSE). Pittsburgh, PA: Carnegie Mellon University, Center for Educational Computing in English.

Norman, D. A. (2002). *The design of everyday things*. New York, NY: Basic Books.

Nozari, N., Freund, M., Breining, B., Rapp, B., & Gordon, B. (2016). Cognitive control during selection and repair in word production. *Language, Cognition and Neuroscience*, *31*, 886-903.

Olive, T. (2004). Working memory in writing: Empirical evidence from the dual-task technique. *European Psychologist*, *9*, 32-42.

Olive, T., Alves, R. A., & Castro, S. L. (2009). Cognitive processes in writing during pause and execution periods. *European Journal of Cognitive Psychology, 21,* 758-785.

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

Ong, W. (2002). *Orality and Literacy.* New York, NY: Routledge

Paivio, A. (1965). Abstractness, imagery, and meaningfulness in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, *4,* 32–38.

Pashler, H. (1993). Doing two things at the same time. *American Scientist, 81*, 48–55.

Pelleg, D., Yom-Tov, E., & Gabrilovich, E. (2015). On the Effect of Human-Computer Interfaces on Language Expression. *arXiv preprint arXiv:1505.00092*.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension Skill. InM. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Oxford, UK: Blackwell.

Peverly, S. T. (2006). The importance of handwriting speed in adult writing. *Developmental Neuropsychology, 29*, 197-216.

Piolat, A., Roussey, J.-Y., Olive, T., & Amada, M. (2004). Processing time and cognitive effort in revision: Effects of error type and working memory capacity. In G. Rijlaarsdam (Series Ed.), L. Allal, L. Chaquoy, & P. Largy (Eds.), *Studies in writing. Revision: Cognitive and instructional processes* (pp. 21-38). Dordrecht, The Netherlands: Kluwer Academic.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879-891.

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*, *16*, 93-115.

Priva, U. C. (2017). Not so fast: Fast speech correlates with lower lexical and structural information. *Cognition*, *160*, 27-34.

Purcell, J. J., Napoliello, E. M., & Eden, G. F. (2011). A combined fMRI study of typed spelling and reading. *Neuroimage, 55*, 750-762.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42,* 107-142.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General, 126,* 211–227.

Rosenbluth, G. S., & Reed, W. M. (1992). The effects of writing-process-based instruction and word processing on remedial and accelerated 11th graders. *Computers in Human Behavior*, *8*, 71-95.

Rumelhart, D. E., & Norman, D. A. (1982). Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science*, *6*, 1-36.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime: User's guide. Psychology Software Incorporated.

Schwarz, N., & Clore, G. L. (1996). Feelings and phenomenal experiences. *Social Psychology: Handbook of Basic Principles*, *2*, 385-407.

Shaw, E. L., Nauman, A. K., & Burson, D. (1994). Comparisons of spontaneous and word processed compositions in elementary classrooms: A three-year study. *Journal of Computing in Childhood Education, 5,* 319-327.

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,* 201-233.

Strijkers, K., & Costa, A. (2016). The cortical dynamics of speaking: Present shortcomings and future avenues. *Language, Cognition and Neuroscience*, *31*, 484-503.

Strömqvist, S. (1999). Production rate profiles. In S. Strömqvist, & E. Ahlsén (Eds), *The process of writing: A progress report* (pp. 53–70). Gothenburg, Sweden: University of Göteborg, Department of Linguistics.

Strömqvist, S., & Ahlsén, E. (Eds.) (1999). *The process of writing: A progress report.* (Gothenburg Papers in Theoretical Linguistics No. 83). Gothenburg, Sweden: University of Göteborg, Department of Linguistics.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*, 24-54.

Templin, M. (1957). *Certain language skills in children: Their development and interrelationships*. Minneapolis, MN: University of Minnesota Press.

Thurstone, L. L. (1938) *Primary mental abilities.* Chicago, IL: University of Chicago Press.

Toglia, M.P. & Battig, W.R. (1978). *Handbook of semantic word norms*. New York, NY: Erlbaum.

Unsworth, N. (2017). Examining the dynamics of strategic search from long-term memory. *Journal of Memory and Language*, *93*, 135-153.

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,* 406-427.

Wengelin, Å. (2007). The word-level focus in text production by adults with reading and writing difficulties. In Rijlaarsdam, G. (Series Ed.); M. Torrance, L. van Waes, & D. Galbraith (Volume Eds.), *Writing and Cognition: Research and Applications* (Studies in Writing, Vol. 20, pp. 67–82). Amsterdam, The Netherlands: Elsevier.

Wagner, R. K., Puranik, C. S., Foorman, B., Foster, E., Wilson, L. G., Tschinkel, E., & Kantor, P. T. (2011). Modeling the development of written language. *Reading and Writing*, *24*, 203-220.

Walker, E. J., Risko, E. F., & Kingstone, A. (2014). Fillers as signals: Evidence from a question–answering paradigm. *Discourse Processes*, *51*, 264-286.

Wershler-Henry, Darren. 2005. *The Iron Whim: A Fragmented History of Typewriting*. Toronto, Ontario: McClelland & Stewart Ltd.

Whittlesea, B. W. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1235-1253.

Winkielman, P., Schwarz, N., Fazendeiro, T, & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K.C. Klauer (eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 189-217). Mahwah, NJ: Erlbaum.

Wolfe, E. W., & Manalo, J. R. (2004). Composition medium comparability in a direct writing assessment of non-native English speakers. *Language Learning & Technology, 8*, 53–65.

Appendices

Appendix A

Narrative prompt (Experiment 1)

Thank you for participating in our study.

Please write an essay on the following topic "*An event that had a positive impact on me*." The essay needs to be at least one (single-spaced) page long. You have 50 minutes to complete the essay. Keep in mind that your essay will be graded, and used in a writing processes study. Therefore, we ask you to imagine that this is an important SAT or GRE assignment.

Before you start writing, please read the following prompts that might help you organize your essay. The purpose of these questions is to help you organize your essay. You do not have to answer the questions, follow the exact order or even use them in your essay.

What has occurred?

What are some words that specifically describe this occurrence?

Where/when did this incident take place?

What are some words that specifically describe the context in which this story took place?

Who was involved in this story?

How did the people involved in this story cause or help to cause it?

What words would you use to describe these people?

What relationship did you have with these people?

Why is this story important to you?

OR

Would you want this incident to occur again? If so, why?

How has this story added to what you now know that you did not know before?

How does what you now know affect your decisions and relations with others?

Appendix B

Narrative prompts (Experiment 2)

B1. Essay 1

Thank you for participating in our study.

Please write an essay on the following topic "*An event that had a positive impact on me (non-school related event)*." The essay needs to be at least half (single-spaced) page long. You have 25 minutes to complete the essay. Keep in mind that your essay will be graded, and used in a writing processes study. Therefore, we ask you to imagine that this is an important SAT or GRE assignment.

Before you start writing, please read the following prompts that might help you organize your essay. The purpose of these questions is to help you organize your essay. You do not have to answer the questions, follow the exact order or even use them in your essay.

What has occurred?

What are some words that specifically describe this occurrence?

Where/when did this incident take place?

What are some words that specifically describe the context in which this story took place?

Who was involved in this story?

How did the people involved in this story cause or help to cause it?

What words would you use to describe these people?

What relationship did you have with these people?

Why is this story important to you?

OR

Would you want this incident to occur again? If so, why?

How has this story added to what you now know that you did not know before?

How does what you now know affect your decisions and relations with others?

B2. Essay 2

Thank you for participating in our study.

Please write an essay on the following topic "*"A memorable school day."* The essay needs to be at least half (single-spaced) page long. You have 25 minutes to complete the essay. Keep in mind that your essay will be graded, and used in a writing processes study. Therefore, we ask you to imagine that this is an important SAT or GRE assignment.

Before you start writing, please read the following prompts that might help you organize your essay. The purpose of these questions is to help you organize your essay. You do not have to answer the questions, follow the exact order or even use them in your essay.

What has occurred?

What are some words that specifically describe this occurrence?

Where/when did this incident take place?

What are some words that specifically describe the context in which this story took place?

Who was involved in this story?

How did the people involved in this story cause or help to cause it?

What words would you use to describe these people?

What relationship did you have with these people?

Why is this story important to you?

OR

Would you want this incident to occur again? If so, why?

How has this story added to what you now know that you did not know before?

How does what you now know affect your decisions and relations with others?

Appendix C

Argumentative prompt (Experiments 3 and 4)

Thank you for participating in our study.

Think carefully about the issue presented in the following excerpt and the assignment below.

Many high school students today regularly carry and use cell phones. Many schools, however, have banned cell phone use on school grounds. Officials at these schools feel that banning cell phones is necessary to create a better environment for learning by eliminating an unnecessary distraction. On the other hand, others feel that cell phones are necessary for scheduling transportation, extracurricular activities, and reaching the police or parents in the event of an emergency. In your opinion, should officials disallow the use of cell phones in schools?

Assignment: In your essay, take a position on the question. You may write about either of the two points of view given, or you may present a different point of view on this question. Use specific reasons and examples to support your position.

The essay needs to be at least one (single-spaced) page long. You have 50 minutes to complete the essay. Keep in mind that your essay will be graded, and used in a writing process study. Therefore, we ask you to imagine that this is an important SAT or GRE assignment.