

Effects of Disfluency in Writing

Srdan Medimorec and Evan F. Risko

Department of Psychology, University of Waterloo, Ontario, Canada

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, we manipulated transcription fluency of ostensibly skilled typewriters by asking them to type essays in two conditions: both-handed and one-handed typewriting. We used the Coh-Metrix text analyzer to investigate the effects of decreased transcription fluency on various aspects of essay writing, such as lexical sophistication, sentence complexity, and cohesion of essays (the important indicators of successful writing). We demonstrate that decreased fluency can benefit certain aspects of writing and discuss potential mechanisms underlying disfluency effects in essay writing.

Effects of Disfluency in Writing

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, Alves, & Castro, 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, Klein, & Brewer, 2014; Diemand-Yauman, Oppenheimer, & Vaughan, 2011). In the current study we 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 RT 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 we 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 deterred 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 (we 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 lesser 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. We test this prediction in the present investigation.

Present Investigation

In the present investigation we examine how typewriting disfluency affects different dimensions of essay writing. In a series of studies we asked participants to typewrite narrative or argumentative essays in a standard way or by using only one hand. Thus, we 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. For example, Matias, MacKenzie, and Buxton (1996) predicted that 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 we are interested in how basic processes in writing are affected by transcription fluency. To address this question we assess the extent to which a battery of text features related to word knowledge, syntactic structure, and cohesion change as a function of our 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 we lose the intuitive appeal of a general quality measure, we gain a more direct and nuanced understanding of how transcription fluency influences writing. In addition, in order to make contact with the previous literature we 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).

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. We 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 supplementary materials). 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 we 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; Vermeer, 2004). 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 are 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. 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 English language. Coh-Metrix computes word frequency using the CELEX database (Baayen, Piepenbrock, & Guilkers, 1995). Here we 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 (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, meaningful, and concrete words are considered more lexically sophisticated (Crosley & 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 are generally more frequent and easier to process (Crossley, Salsbury, & McNamara, 2010).

Sentence complexity

Words before main verb and modifiers per noun phrase. Increased number of words before the 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 score 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.

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, MTL, 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. In addition, we 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. In the majority of cases the two one-handed conditions (i.e., dominant hand vs. non-dominant hand) did not differ. When they did we 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 $F_s < 2.17$, all $p_s > .13$, $d_s < .33$ (see Table 1).

Transcription Fluency

We calculated typewriting transcription fluency as the mean keystroke interval within a word (onset of the current keypress - onset of the previous keypress in ms). We use this measure of typewriting transcription fluency provided within-word keystrokes are only marginally influenced by other writing processes (Strömquist, 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, $F(1,99) = 191.72$, $MSE = 2135.41$, $p < .001$, $d = 2.75$ (see Table 2).

Descriptive Indices

There was no effect of condition on descriptive indices, all $F_s < 1.61$, $p_s > .20$, $d_s < .25$ (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 MTLT $F_s > 4.11$, $p_s < .046$, $d_s > .40$, 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, $F_s > 6.20$, $p_s < .015$, $d_s > .49$. 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 $F_s < 2.09$, $p_s > .15$, $d_s < .29$.

Syntactic Complexity

Condition had a significant effect on the mean number of words before the main verb such that there were less words before 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, $F(1,101) = .19$, $MSE = .02$, $p = .67$, $d = .08$.

Cohesion

Condition did not have a significant effect on cohesion indices, all $F_s < 2.49$, $p_s > .12$, $d_s < .32$ (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, we 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. We 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 we sought to replicate and extend this result in Experiment 2. Specifically, we 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. We 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 order of conditions 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 supplementary materials). One essay was typewritten using both hands, and the other was typewritten by using one hand (dominant or non-dominant). The condition and topic order was counterbalanced. 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 we mention it in the respective analyses.

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 $F_s > 23.43$, $p_s < .001$, $d_s > .36$. 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).

Transcription Fluency

Based on the established criteria we 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 $F(1,66) = 16.14$, $MSE = 1785.28$, $p < .001$, $d = .98$.

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 $F_s > 4.91$, $p_s < .031$, $d_s > .23$. 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 $F_s < 1.22$, $p_s > .27$, $d_s < .18$ (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 $F_s < 2.65$, $p_s > .10$, $d_s < .18$.

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 Experiment 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, we 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, we 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). We 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, $F(1,167) = 8.17$, $MSE = 1812.67$, $p = .005$, $\eta^2 = .047$. Moreover while there was no statistical difference in transcription 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, $F(1,87) = 6.66$, $MSE = 2183.23$, $p = .012$, $d = .59$. We 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 we 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 we attempt to replicate and extend the results of Experiment 1 by returning to a longer essay format. We introduced one critical change in order to assess the generalizability of our results, specifically we changed the essay genre from narrative to argumentative.

Method

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 supplementary materials). 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 $F_s > 4.87$, $p_s < .030$, $d_s > .40$. The other measures were not affected, $F_s < 2.40$, $p_s > .12$, $d_s < .30$ (see Table 5).

Transcription Fluency

Based on the established criteria we 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).

Descriptive Indices

Using all essays, condition had a significant effect on number of words, such that there were more words in the 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, $F_s < 2.58$, $p_s > .11$, $d_s < .30$ (see Table 5).

Lexical Sophistication

Condition had a significant effect on TTR and MTLTD such that the one-handed condition was more lexically diverse compared to the both-handed condition, $F_s > 5.16$, $ps < .026$, $ds > .41$, 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, $F_s > 4.20$, $ps < .044$, $ds > .37$. The other lexical sophistication indices were not affected, $F_s < 1.81$, $ps > .18$, $ds < .25$.

Syntactic Complexity

Condition did not have a significant effect on syntactic complexity indices, $F_s < .12$, $ps > .73$, $ds < .07$. 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 $F_s > 4.45$, $ps < .038$, $ds > .38$. The other cohesion indices were not affected, $F_s < .11$, $ps > .74$, $ds = .06$ (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 benefits 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. Future work directly manipulating the amount of practice (e.g., prior to writing the essay) would further elucidate the mechanisms underlying the differences we 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 our longer essay studies, we next combine data from Experiments 1 and 3 to provide a test of transcription fluency effects on lexical indices on a larger sample. Thus, we analyzed narrative and argumentative essays typewritten by 220 participants (both hands = 111). We 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 a 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). We report significant interactions.

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, $F_s < 1.23$, $p_s > .27$, $d_s < .17$.

Essay type did not have an effect on number of words, $F(1,218) = .13$, $MSE = 29992.21$, $p = .72$, $d = .05$. On the other hand, 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 less paragraphs $F(1,218) = 8.74$, $MSE = 3.37$, $p = .003$, $d = .40$, in argumentative essays compared to narrative essays (see Table 8).

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 $F_s > 5.10$, $p_s < .026$, $d_s > .30$, 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, $F_s > 11.54$, $p_s = .001$, $d_s > .39$. 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, $F_s < 2.55$, $p_s > .11$, $d_s < .21$ (see Table 9).

There was a marginal effect of essay type on MTLT, 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 $F_s > 73.32$, $p_s < .001$, $d_s > 1.15$. On the other hand, 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, $F_s < 2.00$, $p_s > .16$, $d_s < .17$ (see Table 10).

Syntactic Complexity

Condition did not have a significant effect on any syntactic complexity index, $F_s < 2.31$, $p_s > .13$, $d_s < .16$.

On the other hand, 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 the both-handed condition compared to

the one-handed condition, $F_s > 6,71$, $ps = .010$, $ds > .34$. The other indices were not affected, $F_s < .34$, $ps > .56$, $ds < .09$.

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 $F_s > 11.96$, $ps < .001$, $ds > .43$. On the other hand, there were more logical connectives in argumentative essays, $F(1,218) = 20.28$, $MSE = 136.75$, $p < .001$, $d = .62$.

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 our fluency manipulation had a similar effect across genres despite clear changes in various indices as a function of essay genre.

Collinearity

In our next set of analyses we explore the extent to which lexical indices representing lexical diversity, word frequency, and content word overlap are correlated. We 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. While all of the measures were correlated to some extent (see Table 11), only the correlation between MTLN 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 MTLN in the combined analysis, we retained that lexical index in subsequent analyses and excluded content word overlap (note that the results are qualitatively similar when using content word overlap).

Regression Analyses

We 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). We address these questions using a series of regression analyses. In the first step of each regression we entered condition (both hands vs. one hand) as the IV, and one of the lexical indices (i.e., MTLN or word frequency) as the DV. In the next step, number of words was entered as an additional IV. Finally, in step 3 we entered the remaining lexical index (i.e., other than the DV) as an additional IV. In this section we report the semipartial correlations (r_s) as a measure of effect size.

In the first set of regression analyses we use Condition (0 = both hands vs. 1 = one hand) as the predictor, and MTLN as the outcome. There was a significant effect of condition on MTLN, $B = 6.78$ [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 MTLN 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 MTLTLD 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 MTLTLD 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).

Discussion

The combined results from Experiments 1 and 3 support the notion that decreased transcription fluency can influence essay writing (see Table 8). Again, the affected indices were lexical diversity, word frequency, and content word overlap. The regression analyses revealed that the effects of condition on lexical diversity (i.e., MTLTLD) 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 MTLTLD.

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 we 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 we tested the prediction that decreased transcription fluency can have beneficial effects on writing. Our 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. 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.

Implications

Research on transcription fluency underlines 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. The present results add an important dimension to existing discussion about how technology shapes human cognition. Understanding how technology can be customized to meet the needs of individuals of different abilities remains an important goal (e.g., Universal Design for Learning; Rose & Meyer, 2006). Future research aiming at better understanding the mechanisms underlying the interaction between internal and external processes in the context of writing will provide further insight into this particular extended cognitive system.

Conclusion

The current study represents a systematic investigation of disfluency effects in a complex cognitive task. Our 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 exists a rich number of interactions between how we write and what is ultimately written.

References

- 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.
- 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.
- Baayen, R.H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (CD-ROM). Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- 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.
- Brown, C. M. (1988). Comparison of typing and handwriting in “two-finger typists.” *Proceedings of the Human Factors Society, 32*, 381–385.
- 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.

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

- 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: 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: 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.
- 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.
- 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.
- 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: Amsterdam University Press.
- Kellogg, R. T., Olive, T., & Piolat, A. (2007). Verbal, visual, and spatial working memory in written language production. *Acta Psychologica*, 124, 382-397.
- Klein, W. (1994). *Time in language*. London: 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.
- 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.
- 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/>.
- 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.
- 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*, 971-990.
- 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.
- 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.
- 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.
- 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.
- Paivio, A. (1965). Abstractness, imagery, and meaningfulness in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 4, 32-38.
- Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension Skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227-247). Oxford: Blackwell.
- Peeverly, S. T. (2006). The importance of handwriting speed in adult writing. *Developmental Neuropsychology*, 29, 197-216.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107-142.
- Rose, D. H., & Meyer, A. (2006). *A practical reader in universal design for learning*. Cambridge, MA: Harvard Education Press.

- 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.
- 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.
- Strömqvist, S. (1999). Production rate profiles. In S. Strömqvist, & E. Ahlsén (Eds), *The process of writing: A progress report* (pp. 53–70). Gothenburg, Sweden: Department of Linguistics, University of Göteborg.
- 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.
- Toglia, M.P. & Battig, W.R. (1978). *Handbook of semantic word norms*. New York: Erlbaum..
- 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.
- 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.

Footnotes

1. We have also assessed the extent to which within-subject design in Experiment 2 contributed to the lack of an effect of condition on the lexical indices. Here, we conducted an analysis of the Experiment 2 data restricted to the first block of trials (essentially turning the design in 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 $F_s < 2$, $ps > .10$, $ds < .38$). Moreover, there was no effect of condition on syntactic complexity indices (all $F_s = .30$, $ps > .58$, $ds = .12$). 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 $F_s < .76$, $ps > .39$, $ds < .21$ (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.
2. When we considered only the first block of trials (i.e., prior to exposure to the other condition) in Experiment 2, 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 $F_s > 4.77$, $ps < .033$, $ds > .50$.

Table 1.

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

Measure	Both Hands		One Hand		Difference
	M	SD	M	SD	Cohen's d
Essay Descriptives					
Words	641	158	595	179	.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.

Table 2.

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

Measure	Both Hands		One Hand		Difference
	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**
Frequency Content Words	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	.13
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	.08
Aspect Repetition	.78	.09	.77	.09	.09
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. ** $p < .05$. *** $p < .001$.

Table 3.

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

Measure	Both Hands		One Hand		Difference
	M	SD	M	SD	Cohen's d
Essay Descriptives					
Words	403	145	350	95	.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. *** $p < .001$.

Table 4.

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

Measure	Both Hands		One Hand		Difference Cohen's d
	M	SD	M	SD	
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
Frequency Content Words	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. *** $p < .001$.

Table 5.

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

Measure	Both Hands		One Hand		Difference
	M	SD	M	SD	Cohen's d
Essay Descriptives					
Words	644	213	579	132	.38**
Sentences	28.16	9.18	25.95	6.89	.27**
Words per Sentence	23.26	4.51	22.98	5.15	.06
Paragraphs	5.17	1.50	4.75	1.37	.30
TLX					
Mental Demand	-.12	3.90	.66	4.42	.18
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	.04
Effort	.04	4.33	1.78	4.18	.41**
Frustration	-4.47	4.93	-1.64	5.91	.54**

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

Table 6.

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

Measure	Both Hands		One Hand		Difference Cohen's d
	M	SD	M	SD	
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**
Frequency Content Words	2.34	.12	2.29	.11	.41**
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. *** $p < .001$.

Table 7.

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

Measure	Both Hands		One Hand		Difference
	M	SD	M	SD	Cohen's d
Essay Descriptives					
Words	642	188	588	155	.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. ** $p < .05$.

Table 8.

*Means and Standard Deviations of Essay Descriptive Measures Across the Two Essay Types
(Combined Study)*

Measure	Narrative		Argumentative		Difference
	M	SD	M	SD	Cohen's d
Essay Descriptives					
Words	620	169	611	179	.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.

Table 9.

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

Measure	Both Hands		One Hand		Difference Cohen's d
	M	SD	M	SD	
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**
Frequency Content Words	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 Across the Two Essay Types (Combined Study)

Measure	Narrative		Argumentative		Difference
	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***
Frequency Content Words	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. ** $p < .05$. *** $p < .001$.

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. *** $p < .001$.