The effect of automated fluency-focused feedback on text production

Emily Dux Speltz & Evgeny Chukharev-Hudilainen

Iowa State University, Ames, Iowa | USA

Abstract: This article presents a new intervention for improving first-language writing fluency and reports an empirical study investigating the effects of this intervention on process and product measures of writing. The intervention explicitly encourages fluent text production by providing automated real-time feedback to the writer. Participants were twenty native-English-speaking undergraduate students at a large Midwestern university in the United States, all of whom were proficient writers. Each participant composed two texts (one in each of the control and the intervention condition) in an online text editor with embedded keystroke logging capabilities. Quantitative data consisted of product and process measures obtained from texts produced by participants in the control and the intervention condition, and qualitative data included participants' responses to an open-ended questionnaire. Linear mixed-effects regression models were fit to the quantitative data to assess differences between conditions. Findings demonstrated that there were significant differences between the intervention and the control condition in terms of both the product and the process of writing. Specifically, participants wrote more text, expressed more ideas, and produced higher-quality texts in the fluency-focused intervention condition. Qualitative findings from questionnaire responses are also discussed.

Keywords: writing fluency, fluency interventions, product and process measures of writing

journal of WRITING RESEARCH

Dux Speltz, E., & Chukharev-Hudilainen, E. (2021 - accepted for publication). The effect of automated fluency-focused feedback on text production. *Journal of Writing Research, volume(issue),* ##-##. DOI: xx

Contact: Evgeny Chukharev-Hudilainen, Iowa State University, Department of English, 351 Ross Hall, Ames, Iowa 50011 | USA -- evgeny@iastate.edu

1. Introduction

Effective writing skills are crucial for success in education and the workplace, and for integration and participation in an increasingly connected and information-rich society. Most salaried jobs include writing duties (National Commission on Writing, 2004). However, outcome measures show that the teaching of writing has ample room for improvement. For example, in the United States of America, best-available data indicate that only 27% of 12th graders are writing at or above a proficient level – the national policy target – while more than half were rated basic and more than 20% below basic (National Center for Education Statistics, 2012). College instructors estimate that more than half of their students are not prepared for college-level writing (Duncheon & Tierney, 2014).

Learning to write effectively requires many hours of practice (Kellogg, 2008), in part because producing high-quality text can place high demands on writers' cognition. Accounts of mental processes that underlie writing (Hayes, 2012; van Galen, 1991; Olive, 2014), following models of speech (Dell, 1986; Levelt, 1999; Skehan & Foster, 1997), describe text production as a cascade of cognitive processes that start from the writer's intended message and end with keystrokes (or pen strokes). Early, high-level processes in this cascade are responsible for generating ideas and making decisions about appropriate rhetorical strategies and discourse structures. This high-level processing has been described as a conscious, deliberate, and typically effortful "problem-solving" activity (Flower & Hayes, 1980; Hayes & Nash, 1996). Later, low-level processes in the cascade are responsible for lexical retrieval, grammatical encoding, and spelling. In young and developing writers, these processes are also effortful, but with instruction and practice, they become increasingly automatized, allowing attention to be devoted to high-level thinking and reasoning: The writer plans what they are going to say, and this "idea package" is passed on to the low-level processes in the cascade. While these low-level processes are at work turning ideas into language, the writer can, in parallel, attend to deciding on what to say next.

Successful writing, therefore, requires effective coordination between the high-level and the low-level mental processes, which must be done within tight confines of a limited and transient working memory (Berninger et al., 2002). Language processing, including written language production, is fundamentally "now or never": mental representations of ideas are typically fleeting and must be captured in language before they are lost (Christiansen & Chater, 2016). If the low-level processes are easy for a writer, then writing is fluent at the sentence level, and the writer can focus their attention on higher-level issues, such as ideas and rhetoric. However, if the low-level processing is difficult, it may demand cognitive resources that otherwise could be directed to the high-level aspects of writing. As a result, the writer is likely to produce text with poorer content, structure, and argumentation

(McCutchen, 1996; Torrance & Galbraith, 2006; Bourdin & Fayol, 2002). To give an extreme example, a writer who stops in the middle of a sentence to worry about spelling or making verbs agree with nouns may, in a literal sense, forget what they were going to say next.

Fluency (defined as the ability to produce language quickly and without substantial hesitations), according to a consensus in applied linguistics, is one of the three central constructs that capture language learning and knowledge, the other two being accuracy and complexity (Wolfe-Quintero, Inagaki, & Kim, 1998; Pallotti, 2009). As discussed above, a drop in production fluency by a writer signals redistribution of working-memory allocation as low-level processes demand resources that could otherwise be dedicated to the higher-level ideational planning. Thus, the fluency of text production appears to be a desirable target for pedagogical interventions. However, fluency (unlike accuracy and complexity) is practically ignored by the current writing interventions and is measured, at best, across whole writing sessions as the number of words written per minute (Wolfe-Quintero, Inagaki, & Kim, 1998; Van Waes & Leijten, 2015). More generally, since current instructional approaches are based on evaluating texts that are produced by the students, they do not provide learning opportunities that would help diagnose and remediate issues with low-level fluency of writing. The reason for this limitation is that issues in the writing process are not captured in the final text and thus cannot be addressed in the feedback provided to the writer.

At the same time, recent advances in keystroke logging technology (e.g., Leijten & Van Waes, 2013; Vandermeulen, Leijten, & Van Waes, 2020; Van Waes et al., 2021) allow for unobtrusive capture of all moment-by-moment actions taken during the course of text production (i.e., keypresses with associated time-stamps). Keystroke logs can be collected in real time and immediately used to provide fluency-focused feedback to the writers. Notably, the need for exploring new instructional approaches that would directly address students' writing process in addition to traditional product-focused feedback was highlighted a decade ago in a white paper written by leading experts in writing instruction (Graham, Harris, & Herbert, 2011). Until recently, however, technology limitations have precluded such approaches from being developed and widely implemented. The goal of this paper is to take a first step toward filling this gap.

To this end, in the present paper, a fluency-focused intervention is proposed and evaluated. Unlike existing approaches to writing instruction, where the pedagogical focus is set on the improvement of the linguistic accuracy, complexity, or rhetorical effectiveness of writing, in a fluency-focused intervention the pedagogical aim is to help students learn to achieve and maintain high fluency of text production. In line with the above discussion, the writer's pauses may be indicative of suboptimal distribution of their attention between higher-level and lower-level concerns; if that is the case, attaining more fluent output may benefit

the writer and allow them to produce a higher-quality text. On the other hand, pausing may be necessary to allow sufficient time for important decision-making, and then limiting the writer's ability to pause and think might have a detrimental effect on the quality of writing. Arguably, much depends on the writer's proficiency level and the characteristics of the task at hand: A successful writer might pause a lot when composing a complex text but write simpler texts more fluently. Thus, an intervention that encourages writing fluency should be carefully contextualized and grounded in learner needs analysis.

It is important to note that, even though an optimal balance between writing fluently and pausing to think might need to be found by the writer on a task-by-task basis, there may still be a benefit in exposing all writers to the strategy of deliberately fluent text production (Feng & Chukharev-Hudilainen, 2017). Once this strategy is learned, writers can then deploy it when necessitated by the given task. This is similar to other strategies, such as explicit planning or outlining, that are routinely taught in composition courses: Writers might not need to explicitly plan every text that they produce, but having knowledge of and access to a planning strategy is beneficial for the writer.

2. The Present Study

In line with the above discussion, the purpose of the present study is to investigate how writers may react to a fluency-focused intervention, that is, an intervention in which they receive immediate, real-time feedback that encourages them to produce text more fluently (i.e., without pausing). To avoid the above-mentioned controversy related to the potential of a fluency-focused intervention to be useful in some contexts and detrimental in others, the present study focuses on a context where promoting writing fluency is undoubtedly valuable, in and of itself. This is, specifically, the context where writers produce a relatively simple text (i.e., a text that is expected to be at or below their proficiency level), but under an explicit time constraint. Common real-world instantiations of such context include, for example, producing an email or a memo in the workplace with a tight deadline or a large number of competing demands. The present study is situated in the context of college-level academic writing which, due to complex genre expectations, is similar in many ways to language-learning contexts discussed above.

The intervention investigated in the present study comprises automated fluency-focused feedback that is delivered to students in real time, as they type a text on a computer, and is based on the analysis of keystroke logging data. This study compares the intervention to a control condition, where students are informed about the benefits of fluent text production and are aware of the time constraint but do not receive real-time feedback about the fluency of their text production. The present study is driven by the following research questions:

RQ1. What effect does the real-time, automated, fluency-focused feedback have on the process of text production and the produced text?

RQ2. What are the writers' perceptions of such an intervention?

3. Methods

3.1 Participants

The participants in this study were 20 adult (mean age: 20 years, range: 19–23; 17 female), native-English-speaking undergraduate students at a large Midwestern research university in the United States. All participants were skilled writers: They had successfully completed (or tested out of) the two required composition courses at their university and were employed at the university's Writing and Media Center, which provides writing consultancy services to university students. Participants volunteered to take part in this study and were not compensated for their participation.

3.2 Intervention

The proposed intervention aimed to teach the strategy of fluent text production to participants. First, participants were informed about the benefits of writing fluently. Then, they composed texts on a computer under two conditions. In both conditions, participants were given an argumentative prompt and a time limit (10 minutes) to complete a text to the prompt. It was assumed (and confirmed in the pilot tests prior to the study) that participants would be able to produce complete texts to the two prompts within 10 minutes. The imposed time limit was, by itself, encouraging fluent text production.

In the control condition, no additional feedback was provided during the composition process. In the intervention condition, the participant's screen slowly faded once a pause was detected. Specifically, for each second of an inter-keystroke interval during typing, the opacity of the text on the screen was decreased incrementally at the rate of 20% per second; thus, when the inter-keystroke interval reached five seconds, the text on the screen became completely transparent and invisible to the participant (Figure 1). Once the participant resumed typing, the text reappeared in the normal black font. It was explained to the participant that the goal of this manipulation was to provide feedback on their pausing behavior and encourage them to write fluently so as to avoid losing sight of their text. Participants were encouraged to avoid the urge of "overriding" the feedback by typing and removing random characters as such behavior negating the purpose of the study was observed in pilot sessions conducted before the study commenced.

The two argumentative prompts used in this study were as follows: (1) "Should there be free college for everyone in the United States? Present both sides of the

argument." and (2) "Should Americans have the right to buy guns? Present both sides of the argument." These prompts were selected because they encouraged participants to utilize higher-order writing skills (such as developing claims, incorporating evidence, and organizing ideas) to develop both sides of an argument; at the same time, the topics were accessible for and well-known to college students in the United States, thereby not requiring them to conduct outside research in order to complete the tasks.

This study utilized CyWrite (Chukharev-Hudilainen et al., 2019; Feng et al., 2016; Chukharev-Hudilainen & Saricaoglu, 2016), a web-based tool for research into writing processes that features a user interface providing a familiar word processing experience (i.e., similar to Microsoft Word or Google Docs). CyWrite has built-in functionality to capture writing-process data as a keystroke log that can then be analyzed both via visualized writing-session playbacks and quantitatively, by exporting variables of interest as a machine-readable dataset. For the purpose of this study, the authors modified the code of CyWrite so as to enable the fading of the text in the intervention condition.

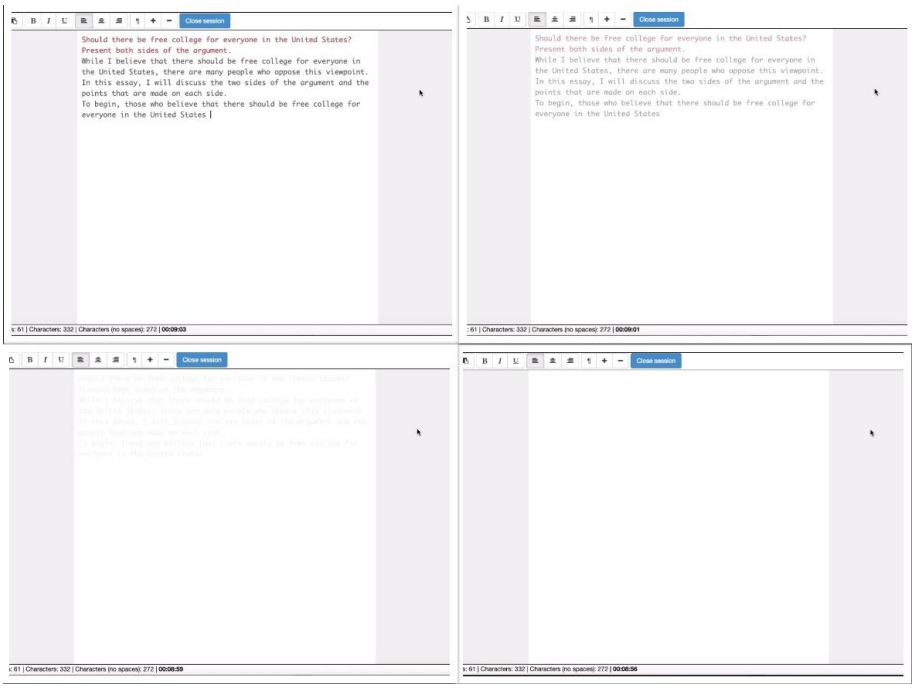

Figure 1. Fading of text in the intervention condition from 100% opacity to 0% opacity upon pausing.

3.3 Procedures

Upon informed consent, each participant met one-on-one with the first author of the present paper for a single data collection session that lasted less than 1 hour. The study and its purpose were explained in full, including the tasks and procedures, at the beginning of the meeting. The benefits of writing fluently were also explained. The researcher showed the participants what they would be doing on the CyWrite website and answered any questions that they had about the task. The study followed a counterbalanced within-participant design, with counter-balancing for prompt order and condition order (control vs. intervention). Each participant was randomly assigned to one of the four counter-balancing groups (Figure 2).

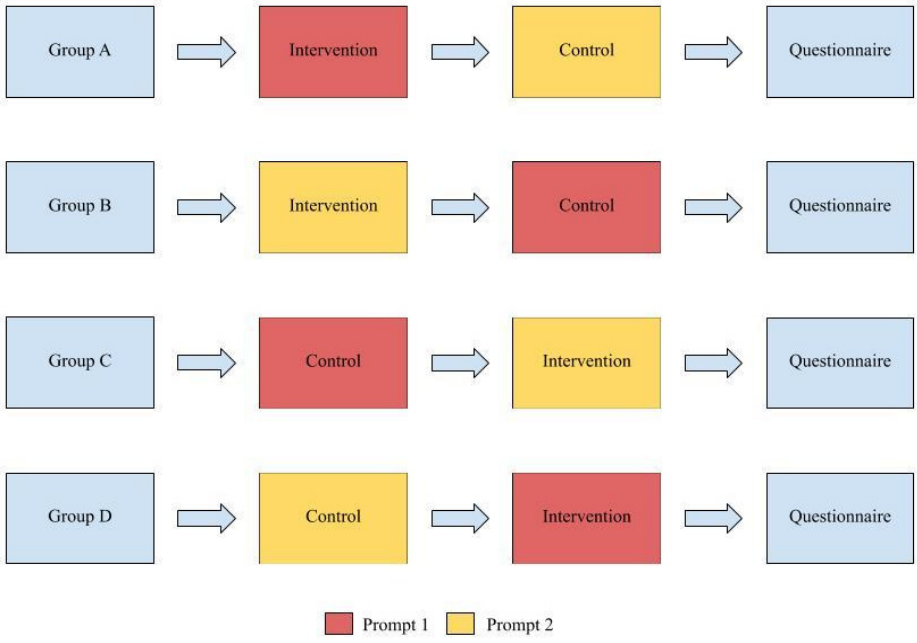

Figure 2. Counter-balancing groups in the study.

Participants first wrote for ten minutes in response to the first prompt that they were given according to their group assignment. At the end of the ten minutes, the CyWrite system automatically stopped participants' composition sessions. They were allowed to take up to ten minutes for a break between tasks. Then, they proceeded to write to the second assigned prompt, again for ten minutes. Similarly, the CyWrite system automatically stopped the task after those ten minutes expired.

After participants were finished with both writing tasks, they were asked to respond to a questionnaire about their experience with the intervention by writing their responses on a piece of paper; this method, as opposed to oral post-session interviews, was used to receive more honest responses since participants did not have to explain their opinions orally to the researcher. In the four questions, the intervention condition was called "the disappearing text condition" to streamline participants' understanding. The questions were formulated as follows:

1. How did you feel as you were writing with the disappearing text condition?
2. Did you feel like you had to sacrifice grammatical/spelling accuracy while writing with the disappearing text condition?
3. Did you feel like you had to sacrifice complexity while writing with the disappearing text condition?
4. How do you think the disappearing text condition impacted your writing process?

3.4 Measures

To answer RQ1, the dependent variables used in this study captured both the writing process and the written product. Process indices included text production fluency, measured as the number of characters typed per minute (both including and excluding subsequently deleted characters), and the percentage of characters deleted during the writing process. Additionally, distributions of inter-keystroke intervals (IKIs) at linguistically-relevant locations in the text (sentence-initial, word-initial, and mid-word) were analyzed. These measures were automatically extracted from the CyWrite keystroke log files.

Written-product measures utilized in this study included measures of complexity, accuracy, discourse density (i.e., the number of ideas expressed), and holistic text quality.

Complexity measures in this study focused on grammatical complexity. While including lexical complexity measures might also be warranted, the number of measures tested needed to be limited so as to alleviate the multiple testing problem. There were two sets of grammatical complexity measures: T-unit-based and phrase-based. The former set included the number of T-units (units consisting of an independent clause and all of its subordinate clauses) per text and the mean length of T-units. For example, the following sentence contains a single T-unit with multiple embedded subordinate clauses: "Education offers an avenue [to pursue the career paths [that intrigue a person]]" (Participant 6, Prompt 1). T-unit-based measures are commonly used in research studies investigating grammatical complexity. In addition, two phrase-based complexity indices (the number of attributive adjectives and the number of premodifying nouns) were utilized based on the argument by Biber et al. (2011) that these capture syntactic complexity in academic writing better than traditional, T-unit-based measures. The number of T-

units per text was manually calculated by the first author. A second annotator (a graduate student in applied linguistics) independently repeated the task on a random subset of 20% of the texts (n = 8). Inter-annotator reliability was assessed using Krippendorff's α with the interval metric (Hayes & Krippendorff, 2007; Krippendorff, 2007), yielding $\alpha_{interval} = 1.0$ (perfect agreement). Mean length of T-unit was then calculated by dividing the length of text in words by the number of T-units in the text. For phrase-level measures, texts were tagged with the part-of-speech tagging software CLAWS (Rayson & Garside, 1998), and then a script was used to count the features of interest based on part-of-speech tag sequences. Counts were normed per 100 words.

Linguistic accuracy was measured by calculating the number of error-free T-units, the rate of error-free T-units per total T-units, and the rate of errors per T-unit. These were calculated manually by the first author, and the procedure was repeated independently by a second annotator, a graduate student in applied linguistics, on the entire corpus. Inter-annotator reliability was perfect, $\alpha_{interval} = 0.99$.

Next, the first author manually marked discourse constituency units (DCUs) in each text and calculated the number of DCUs per text. DCUs, "atomic units of discourse" (Polanyi, 1995, p. 2), were defined as "a contextually indexed representation of information conveyed by a semiotic gesture, asserting a single state of affairs or partial state of affairs in a discourse world" (p. 5). Operationally, a DCU was identified when sentential syntactic or semantic criteria indicated a change in any of the contexts (spatial, temporal, modal, etc.) that index the discourse worlds where the events (or states of affairs) are interpreted. For example, the following sentence contains six DCUs, indicated by brackets: "[Proponents of free college argue] [that the high cost of college] [ostracizes poorer individuals] [with potential for success] [and that graduates are being crushed] [under the costs of student debt and tuition]" (Participant 9, Prompt 1). The number of DCUs, then, was used as a measure of discourse density, or the number of ideas expressed in the text. The second author independently annotated DCUs for a random subsample of 20% of the texts (n = 8), yielding $\alpha_{interval} = 0.98$ (perfect agreement).

Finally, a holistic measure of overall writing quality was obtained from two independent raters (both graduate students in applied linguistics) using a holistic writing rubric (see Appendix). The rubric was based on one used in first-year written communication courses that every participant in this study had completed, but band descriptors were modified to be task-specific. The rating was done on a scale from 1 to 5 with a step of 0.5 (a total of 9 discrete points). To calculate inter-rater reliability for these measures, both raters evaluated an overlapping random subsample of 15% of the texts (n = 6). Krippendorff's α showed high reliability at $\alpha_{interval} = 0.92$. For the texts in the reliability subsample that received different scores from the raters, the average score was used for further analysis.

3.5 Statistical Analysis

To address the first research question, a series of linear mixed-effects regression models were run using the lme4 package in R. In each model, the intercept was allowed to vary by participant as a random effect. For each dependent variable of interest, the data were first screened by plotting a histogram. If the data visibly deviated from the normal distribution, they were log-transformed. Log-transformations yielded acceptable results in all cases. Then, for each dependent variable, two nested models were built: the baseline, intercept-only model; and the model adding a fixed effect for Condition (control vs. intervention). The likelihood-ratio test was used to evaluate the gains in the goodness of fit of the second model relative to the first one. If the full model (i.e., the one adding the fixed effect for Condition) fit the data significantly better than the intercept-only model, it was then concluded that the Condition significantly affected the dependent variable of interest. Wald estimates of the confidence intervals (CIs) for means of the dependent variables were then derived from the full models.

The analysis of inter-key intervals (IKIs) followed the methodology proposed by Chukharev-Hudilainen and colleagues (2019). The durations of IKIs during text production were extracted from the keystroke logs provided by the CyWrite system. Using a script, these IKIs were automatically classified into sentence-initial, word-initial, and within-word. A sentence-initial IKI was defined as occurring before typing a capital letter after a sentence-final punctuation character (a period, a question mark, or an exclamation point) and a space. A word-initial IKI was defined as occurring before typing the first character of a word that was not sentence-initial. Finally, a within-word IKI was defined as occurring before typing a letter that was not word-initial or sentence-initial. Only IKIs produced without an intervening revision were considered. That is, for example, no revision or cursor movement could be initiated between typing the sentence-final punctuation character and the sentence-initial letter. IKI durations were trimmed at 8,000 ms and log-transformed. Linear mixed-effects models were fit to IKI data to predict the duration of IKIs based on the Location (i.e., sentence-initial, word-initial, within-word) and the Condition. Thus, four nested models were built: an intercept-only model; a model adding a fixed main effect of Condition; a model adding a fixed main effect of Location; and a model adding an interaction between Condition and Location. Again, the likelihood-ratio test was used to evaluate the gains in the goodness of fit of each subsequent model relative to the previous one; a significant gain indicated that the fixed effect added to the subsequent model was a significant predictor of the dependent variable. Wald estimates of the means and CIs of IKI durations at each Location and in each Condition were then derived from the full model.

To address RQ2, simple descriptive statistics were utilized to explore the frequency of participant responses that fell under inductively derived thematic categories. The themes that emerged from the analysis of participant responses

were coded by the first author. However, it was not necessary to employ a second coder because all participants used the same predictable keywords to refer to each emotion, and these keywords could be formally extracted from the participants' responses to the questionnaire.

4. Results and Discussion

Participants produced texts that were, on average, 299 words long (range: 111–474; *SD* = 98.4). None of the participants completed the assignment ahead of time, so the duration of each writing session was exactly 10 minutes. Below, we discuss findings in terms of process and product measures obtained from the texts.

4.1 Process Measures

Not unexpectedly, process measures showed a high degree of correlation as they all captured the same construct of text-production fluency (Table 2).

Table 2. Correlation matrix for process measures

	Characters per Minute (including deletions)	Characters per Minute (excluding deletions)	Percentage of Characters Deleted
Characters per Minute (Process)	1		
Characters per Minute (Product)	0.95	1	
Percentage of Characters Deleted	-0.56	-0.78	1

Table 3 shows the means and 95% CIs of the three process measures for both conditions, along with results of significance tests.

As it can be seen from the table, participants wrote significantly faster and produced more characters per minute in the intervention condition, both in terms of the total number of characters produced and the number of characters that were kept after deletions in the final product. In terms of the percentage of characters deleted during the writing process, no statistically significant differences were found between conditions. This suggests that participants were not simply writing more characters per minute in the intervention condition for the sake of getting something down on the screen that they could delete later to "override" the fluency-focused intervention. Rather, the lack of a significant difference between conditions for the percentage of characters deleted may suggest that participants took the fluency-focused feedback seriously by producing meaningful text.

Table 3. Process measures across conditions for all participants

	Intervention, Mean [95% CI]	Control, Mean [95% CI]	Significance
Characters per Minute (including deletions)	244 [223, 265]	203 [182, 224]	$\chi^2(1) = 31.60$, $p < 0.001$
Characters per Minute (excluding deletions)	195 [172, 218]	158 [135, 181]	$\chi^2(1) = 24.60$, $p < 0.001$
Percentage of Characters Deleted	0.20 [0.15, 0.25]	0.22 [0.17, 0.27]	$\chi^2(1) = 1.46$, $p = 0.23$

Table 4. Mean IKIs across conditions and locations

	Intervention, Mean [95% CI]	Control, Mean [95% CI]
sentence-initial	320 [288, 356]	488 [437, 545]
word-initial	209 [193, 226]	216 [200, 233]
within-word	122 [113, 132]	123 [114, 132]

Note: See text for the discussion of significance.

The model with one fixed factor of Condition did not fit the data significantly better than the intercept-only model ($\chi^2(1)=0.93$, $p=0.34$), meaning that the main effect of Condition on IKIs was not significant; in other words, overall, the latencies before non-revision keystrokes were not significantly different across the two conditions. However, adding the fixed factor of Location significantly improved model fit ($\chi^2(2)=9095$, $p<0.00$), meaning that the main effect of Location was significant. Finally, adding the interaction between Location and Condition has further significantly improved model fit ($\chi^2(2)=57.27$, $p<0.00$). As can be seen from Table 4, the confidence intervals for the two conditions are almost identical for within-word IKIs, and largely overlap for word-initial IKIs (with participants being faster at the start of the word by only 7 ms in the intervention condition). This allows us to conclude that there was no significant difference between the intervention and the control condition in the duration of IKIs at these locations. However, for sentence-initial IKIs, the confidence intervals for the two conditions do not overlap. Participants were faster at the start of a new sentence by 168 ms in the intervention condition relative to control, and that difference was significant.

4.2 Product Measures

The correlations among product measures are presented in Table 5. Some of the measures showed a high degree of correlation (e.g., the total number of errors per text strongly correlated with the number of errors per T-unit, not surprisingly), while others only showed a low to moderate degree of correlation (e.g., the number of premodifying nouns per text correlated with the number of attributive adjectives very weakly).

Table 5. Correlation matrix for product measures

	T-Units	Words per T-unit	Attributive Adjectives	Premodifying Nouns	Error-free T-unit ratios	Errors per T-unit	Total Errors	DCUs	Holistic Ratings
T-units	1								
Words per T-unit	-0.64	1							
Attributive Adjectives	-0.24	0.29	1						
Premodifying Nouns	0.22	-0.04	-0.05	1					
Error-free T-unit ratios	-0.20	0.34	0.51	-0.06	1				
Errors per T-unit	0.12	-0.29	-0.44	0.08	-0.93	1			
Total Errors	0.59	-0.57	-0.39	0.24	-0.74	0.79	1		
DCUs	0.71	-0.16	-0.03	0.09	0.11	-0.20	0.16	1	
Holistic Ratings	0.13	0.34	0.21	0.21	0.25	-0.31	-0.14	0.34	1

Table 6 presents the means and 95% CIs for the product measures and the significance of the differences between the two conditions. Number of T-units per text, number of words per T-unit, normalized frequency of attributive adjectives, and normalized frequency of premodifying nouns were used as measures of complexity. Rate of error-free T-units, number of errors per T-unit, and total number

of errors were measures of accuracy. The number of DCUs and the holistic ratings of writing quality are also provided as product measures.

Table 6. Product measures across conditions for all participants

	Intervention, Mean [95% CI]	Control, Mean [95% CI]	Significance
T-units	18.5 [15.0, 22.0]	13.5 [11.0, 16.0]	$\chi^2(1) = 11.90$, $p < 0.001$
Words per T-unit	18.5 [17.0, 20.0]	19.0 [17.0, 21.0]	$\chi^2(1) = 0.11$, $p = 0.74$
Attributive Adjectives	4.00 [3.0, 5.0]	4.85 [3.70, 6.00]	$\chi^2(1) = 3.19$, $p = 0.07$
Premodifying Nouns	2.05 [1.40, 2.70]	1.75 [1.20, 2.30]	$\chi^2(1) = 0.57$, $p = 0.45$
Error-free T-unit ratios	0.54 [0.46, 0.62]	0.64 [0.55, 0.72]	$\chi^2(1) = 6.25$, $p = 0.01$
Errors per T-unit	2.00 [1.60, 2.40]	1.75 [1.40, 2.10]	$\chi^2(1) = 8.08$, $p < 0.005$
Total Errors	10.20 [7.40, 13.00]	6.35 [4.50, 8.20]	$\chi^2(1) = 16.30$, $p < 0.001$
DCUs	41.50 [36.0, 47.0]	33.5 [28.0, 39.0]	$\chi^2(1) = 21.00$, $p < 0.001$
Holistic Ratings	3.55 [3.00, 4.10]	3.20 [2.70, 3.70]	$\chi^2(1) = 5.28$, $p = 0.02$

T-unit-based measures of complexity included the number of T-units per text and the number of words per T-unit. In the intervention condition, participant texts contained significantly more T-units than those in the control condition. There was no significant difference between conditions for the number of words per T-unit. As discussed above, these T-unit measures may provide some insight regarding the complexity of the texts, but a more adequate understanding of complexity may be gained by investigating phrasal measures of complexity such as the use of attributive adjectives and nouns as premodifiers in noun phrases.

As seen in Table 6, there was no significant difference between conditions for the number of attributive adjectives and nouns used as premodifiers. Whereas texts in the intervention condition had slightly fewer attributive adjectives, they also had slightly more nouns as premodifiers; however, neither of these differences were statistically significant. From these results, it can be inferred that the fluency-focused intervention did not significantly impact the levels of grammatical complexity present in these texts.

Accuracy was measured in three ways: the ratio of error-free T-units, the number of errors per T-unit, and the total number of errors per text. As shown in Table 6, the ratio of error-free T-units was lower in the intervention condition; this was statistically significant. The number of errors per T-unit and the total number of errors per text were both greater in the intervention condition. In sum, participants wrote with less accuracy in the intervention condition compared to the control.
The number of DCUs was significantly higher in the intervention condition. This demonstrates that participants developed and expressed more ideas when the fluency-focused feedback was provided than in the control condition. There is also some evidence ($p = 0.02$) that holistic writing quality ratings were significantly higher in the intervention condition, implying that not only were the additional ideas beneficial for increasing the word count, but these ideas also helped the participants develop more complete arguments to satisfy the holistic rating requirements.

To further investigate the effect of the intervention on final text quality, the following post-hoc analysis was conducted: An additional pair of nested mixed-effects models were fit to the data, predicting the holistic ratings, but this time adding the total number of characters in the final product (i.e., text length) as a controlling variable. In this series of models, the difference between conditions was not significant: $\chi^2(1) = 0.34$, $p = 0.56$. This suggests that aspects of writing quality that were not directly connected with the length of the text produced (such as, for example, the quality of transitions between paragraphs, etc.) were not significantly improved in the intervention condition relative to control.

4.3 Questionnaire Responses

Questionnaire responses provided information about students' perceptions of the study tasks. They also allowed for a comparison between how participants felt that they completed the tasks and how they actually did. This section will discuss the participants' responses to each of the four questions.

Q1: How did you feel as you were writing with the disappearing text condition?
Participants expressed a wide range of emotions that centered around the main themes summarized in Figure 3.

As Figure 3 shows, the most common feeling evoked by the intervention condition by far was "stressed." It should be noted that the emotions represented in this coding scheme were not mutually exclusive; the total number of responses exceeds 20 because many responses included multiple emotions, and a few even expressed how the participant felt stressed at first but eventually became more focused or comfortable over time. One response even discussed how the initial stress of the task eventually gave way to relief: "I was certainly a bit more stressed out, but it was

also almost a relief to just get all my thoughts and ideas out of my head. I felt more focused on the prompt and task at hand because of the urgency." This type of response suggests the pedagogical potential of the proposed intervention. If students in writing classes are struggling to formulate ideas or get them down in writing, this task could provide a "relief" to such students by forcing them to write their ideas without worrying as much about typical writing conventions, such as proper grammar or organization, for example. It is unclear whether the stress felt by the participants would hinder the use of the intervention for longer tasks, or whether it might be diminished with further practice; further research is needed to better understand how the participants' stress level might change longitudinally when this type of intervention is deployed.

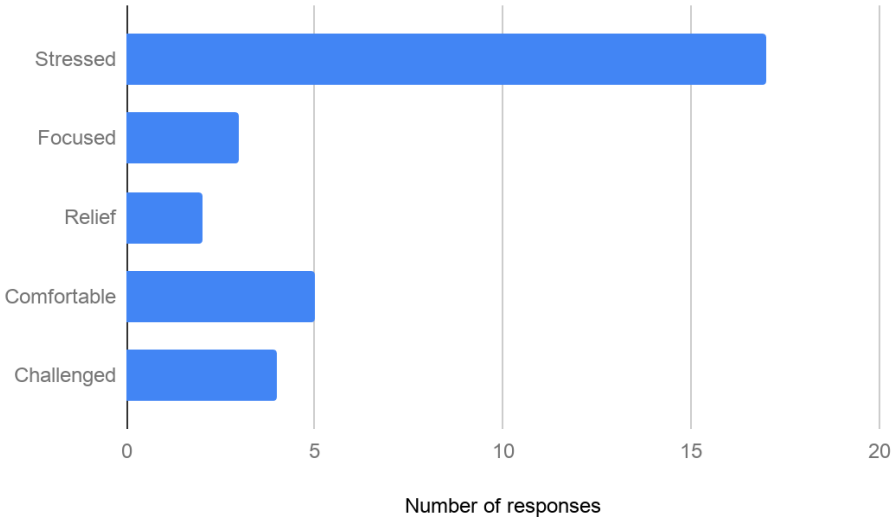

Figure 3. Feelings expressed in participants' responses to "How did you feel as you were writing with the disappearing text condition?"
(Note: Categories were not mutually exclusive.)

Another response expressed that the intervention task was challenging, but it was also enjoyable to approach a writing challenge in a new way: "I also enjoyed how it challenged me to think on my toes and be flexible with my writing and approach." Clearly, this participant recognized a salient benefit of the intervention: More ideas can be generated in this way by forcing participants to think quickly and not delay to hesitate about lower-order concerns.

Q2: Did you feel like you had to sacrifice grammatical/spelling accuracy while writing with the disappearing text condition? A majority of the participants (n = 14; 70%)

reported that they felt they had to sacrifice the accuracy of their grammar and spelling in the intervention condition.

A few participants gave additional comments to explain the reasoning for their answers. For example, one participant noted, "Yes. For more complex concerns, I didn't have time to correct them. I focused on easier things like spelling." Another participant shared the following: "I felt like I had to sacrifice accuracy just in terms of my limited knowledge of the issues. I didn't have time to dwell on whether I accurately communicated the complex ideas—but that is a revising step anyway!" These responses may indicate that participants had different understandings of what "accuracy" meant in the context of their writing. While the former participant is talking about accuracy in terms of word- and sentence-level mistakes, the latter one approaches it from the standpoint of validity of the information that was being conveyed (even though the question explicitly stated "grammatical/spelling accuracy"). This type of misunderstanding may have impacted other participants' responses to the question as well.

Q3: Did you feel like you had to sacrifice complexity while writing with the disappearing text condition? In contrast to the previous question, 55% of participants (n = 11) reported that they did not feel that they had to sacrifice complexity in the intervention condition.

These responses are inconsistent with the results obtained from the complexity measures above since 45% of participants (n = 9) felt that they were sacrificing the complexity of their writing when, in reality, there was no significant difference between the intervention and control conditions. Once again, many participants expressed the reasoning behind their answers, and these comments are worth discussing.

One participant shared why she felt that her complexity was sacrificed: "My sentence structure and word choice were less sophisticated because I was more concerned with writing than with my prose and technical presentation." Another participant shared a different understanding of complexity: "I believe my sentences were still as complex, but the organization of my ideas was not complex at all. Additionally, I usually look up synonyms to diversify my vocabulary, but I did not have time to do that here." Whereas the first comment understood complexity as "sophisticated" sentence structure and word choice, the second comment discussed the organization of ideas and diversity of vocabulary selection. As with accuracy, these varied understandings of complexity may have impacted the answers to this question. Regardless, to combat the negative feelings that complexity had been sacrificed, in potential applications of a fluency-focused intervention, data from complexity measures may be shared with students.

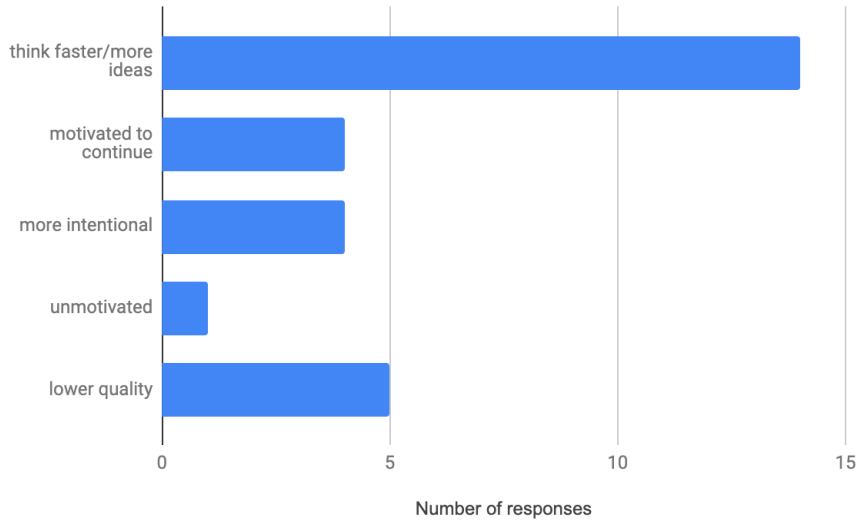

Figure 4. Questionnaire responses to "How do you think the disappearing text condition impacted your writing process?" (Note: Categories were not mutually exclusive.)

Q4: How do you think the disappearing text condition impacted your writing process? This question yielded the most in-depth, and perhaps the most revealing, responses from participants. Figure 4 provides an overview of the general themes that were expressed by the responses to Q4. The total number of responses indicating each theme exceeds 20 because some responses indicated multiple themes.

Although Q1 revealed that a large majority of participants felt that they were stressed while completing the intervention, many responses to this question revealed that participants actually appreciated the changes that happened to their writing process in this condition. Answer 1 (A1) is one such response from the "think faster/more ideas" category:

> A1: "I think it pushed me to consider more avenues around the same prompt. It encouraged me to just get my ideas out instead of truly considering whether they were good or bad. It contributed to stress but resulted a [*sic*] simple rough draft that would generally take me longer as I would want it to be more polished."

In this response, the participant acknowledged the stress, but she also appreciated the way she was able to "consider more avenues" and produce a draft in less time than she would normally take. A2, an answer from the "more intentional" category, also presents valuable benefits to the intervention condition:

> A2: "I actually liked the disappearing text more. I felt like it forced me to be more intentional in what I wrote because I didn't have time to read back. For the prompt, I felt strongly one way, but the condition made me quickly consider the other side. I actually wrote the opposing viewpoint first, which I don't think I would have done otherwise. Overall, I felt that this essay was the stronger of the two."

A2 presents an opposing view to A1 in that the participant felt that she had to be more intentional in her writing, whereas the A1 participant felt that she had to create a rough draft that was less polished than she would normally write. Interestingly, despite these two opposing viewpoints, both participants found the strategy to be beneficial to their final product. Additionally, A2 shows that the participant felt that the intervention allowed her to consider more about the opposing viewpoint from her own because she had to quickly consider more ideas to discuss to continue writing without excessively pausing. Since part of the requirement of the prompt included a discussion of both sides of the argument, this was beneficial for this participant's final product.

A3 provides an example of a response that fits into the "think faster/more ideas" and the "more intentional" categories. It also captures one of the potential ways that a fluency-focused intervention could be used in teaching writing:

> A3: "It helped me start simply getting my ideas onto the page, which is what I struggle most with in my writing process. I also have a tendency to revise heavily while I write, so I felt like I could really focus on the actual content with the forced fluency instead of getting hung up on writing style. This would be hugely helpful in the drafting stage."

This participant recognized the way that a fluency-focused intervention could potentially be used in the drafting stage of composition in order to allow writers to focus more on content and less on writing style. Additionally, this participant also recognized the clear benefit of increased idea production.

A4 builds on this idea of increased idea production ("think faster/more ideas") by recognizing the way that this participant was able to make connections:

> A4: "At one point, I definitely felt like I made a connection I wouldn't have without it (as of desperation to keep the text there). That was cool to see. I think my overall quality was similar to the 'normal' one, and I definitely wrote more."

The "desperation" to prevent the text from fading during the fluency-focused intervention allowed this participant to not only produce more ideas but also connect her ideas in an unforeseen way. This benefit should not be overlooked when considering the applications of this type of intervention in future writing contexts.

A5 provides an insightful (and almost humorous) evaluation of this continued idea of increased idea production in the intervention task:

> A5: "I didn't know if any of my ideas were worth anything, but potentially losing those ideas freaked me out."

Although the goal of this intervention was not to "freak out" the participants, it seems to have been successful in allowing participants to develop ideas first and then consider those ideas' worth after they were documented in writing. This answer was classified in the "motivated to continue" category as it also demonstrates how the intervention motivated this participant to continue writing despite hesitation about the quality of their ideas.

In sum, these questionnaire responses provided insight into RQ2 by showing participants' perceptions of the intervention. While the majority of the participants reported feeling stressed from the intervention condition, they also recognized its potential to keep them focused and to challenge them as writers. Most participants recognized that their accuracy was sacrificed by writing in this condition, but they did not feel that they had to sacrifice the complexity of their writing. This finding is promising for the potential real-life use of this intervention as writers can be told that time for revision will be allowed later. Therefore, writers may feel less concerned about the intervention's impact on their accuracy if they know they do not have to be perfectly accurate on their first draft during the intervention. Finally, participants identified many benefits of the intervention on their writing process, specifically citing thinking faster, generating more ideas, being motivated in their writing, and being more intentional in their work. Some participants expressed that they felt their writing was of lower quality, but ideally, reporting that holistic quality ratings were significantly higher during the intervention may relieve future users of this concern.

5. Conclusion

This study explored the potential of a computer-assisted intervention that directly encouraged fluent text production through automated real-time feedback to the writer. The study was guided by two research questions: (1) What effect does the real-time, automated, fluency-focused feedback have on the process of text production and the produced text? and (2) What are the writers' perceptions of such an intervention?

The first question was addressed by comparing product and process measures of participants' writing between the intervention and the control condition. It was found that although accuracy suffered and complexity remained unchanged in the intervention condition, all other measures significantly improved, demonstrating that this intervention was not only successful at improving fluency, but also has the

potential to improve the writing outcomes. Holistic quality gains between the two conditions, however, were no longer significantly different when the length of the final text produced by the writers was controlled for. The significant effect of final text length on writing quality was expected because producing a more sophisticated argument requires writing more text. At the same time, the lack of quality difference after controlling for length suggests those aspects of writing quality that might not be directly associated with the length of the text (such as, for example, coherence) were not different across the two conditions.

An analysis of IKIs revealed that the attainment of higher fluency was primarily through the reduction of pauses that occurred at the start of sentences. Word-initial pauses and mid-word pauses were not significantly different between the two conditions. Thus, reducing the amount of cognitive processing at the beginning of each sentence has allowed the writers to produce longer (and better) texts, albeit less accurate. Since keystroke latencies are explained, in part, by planning the linguistic unit that follows, it can be inferred that (1) in the control condition, participants may have been "overthinking" their ideas or "overplanning" the structure of their sentences, and the fluency-focused intervention may have boosted their productivity by reducing the amount of such "unnecessary" planning; and (2) linguistic decisions that are important for attaining accurate output may happen at the start of a sentence to a larger extent than mid-sentence. Both of these inferences call for additional exploration in focused follow-up studies.

The second question was addressed by administering a questionnaire to participants after they completed both writing tasks. While participants revealed that the intervention condition made them feel stressed, they also identified many benefits of this condition. They recognized their ability to think faster and generate more ideas, and some participants even reported feeling more focused, intentional, and motivated in their writing. If a fluency-focused intervention is implemented with students of academic writing, such students could be proactively informed about the findings of the present study in terms of the intervention's positive impact on fluency, ideational density, and holistic quality to both lessen their stress about the intervention as a whole and, ideally, to alleviate concerns about lower-level issues such as accuracy. It could be emphasized to future users of this intervention that revisions can be made to improve accuracy after the initial writing session with fluency-focused feedback.

Several limitations were present in this study. First, the study did not collect eye movement data which could be useful in interpreting the findings of the present study. Knowing where participants looked during sentence-initial pauses that were different across conditions could provide further insight into the differences in cognitive processing that occurred during these pauses. Future studies, therefore, could incorporate eye-tracking technology.

Second, participants only composed texts that were below their proficiency levels (as they had successfully completed classes that taught this kind of writing). While this was a deliberate choice that aided in the exploration of the benefits of forced fluency under optimal conditions, it is not clear how a similar intervention would affect writing that occurs at or beyond the writer's level of proficiency (for example, when composing more complex discipline-specific texts, writing from sources, or writing in a non-native language). In addition, the very short duration of the writing session might have signaled "a drafting task" to the participants, thus discouraging revision behavior that would otherwise be expected.

Third, future work is needed to investigate the relationship between editing, time pressure, and accuracy. The current study has found that the accuracy in the intervention condition was lower than in the control condition, and many of the research participants themselves noted that they were forced to sacrifice the linguistic accuracy of their writing. At the same time, the percentage of the characters deleted by participants did not significantly vary across conditions; therefore, the sacrifice of accuracy was not made through the reduction in editing. A future study, then, might investigate participants' editing behavior in more detail in terms of the types (low-level vs. high-level) and the outcomes (successful vs. unsuccessful) of revisions that they made in the two conditions.

Despite its limitations, this study paves the way for developing fluency-focused interventions that could be useful in writing pedagogy. The study demonstrated that the real-time fluency-focused feedback was useful not only for increasing the fluency of written text production, but also for improving the overall writing quality. Therefore, future studies may investigate the benefits of this type of intervention for writing pedagogy (for example, in the context of writing classrooms or university writing centers). The fluency-focused intervention could also be used to specifically target students who may struggle to write fluently. For example, if a student is pausing too frequently or for too long (and such pausing behavior is diagnosed), then the student might benefit from an intervention similar to the one described in the present article. On the other hand, some students might not benefit from this type of intervention, for example, if they are already sufficiently fluent in their writing process and might not need encouragement to produce their text even faster. Future work should focus on implementing and evaluating such pedagogical interventions, as well as on investigating their effects on the long-term learning gains in the learners.

Acknowledgements

This material is based upon work supported by Iowa State University Department of English through a Departmental Research Grant Award to the first author and by the National Science Foundation under Grant No. 2016868 to the second author.

The authors are grateful to the anonymous reviewers for the valuable feedback on early versions of this manuscript.

References

Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., Hawkins, J. M., & Graham, S. (2002). Teaching spelling and composition alone and together: Implications for the simple view of writing. *Journal of Educational Psychology, 94*(2), 291–304. https://doi.org/10.1037/0022-0663.94.2.291

Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., Hawkins, J. M., & Graham, S. (2002). Teaching spelling and composition alone and together: Implications for the simple view of writing. *Journal of Educational Psychology, 94*(2), 291–304. https://doi.org/10.1037/0022-0663.94.2.291

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *Tesol Quarterly, 45*(1), 5-35. https://doi.org/10.5054/tq.2011.244483

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences, 39.* https://doi.org/10.1017/S0140525X1500031X

Chukharev-Hudilainen, E., & Saricaoglu, A. (2016). Causal discourse analyzer: Improving automated feedback on academic ESL writing. *Computer Assisted Language Learning, 29*(3), 494-516. http://dx.doi.org/10.1080/09588221.2014.991795

Chukharev-Hudilainen, E., Saricaoglu, A., Torrance, M., & Feng, H. H. (2019). Combined deployable keystroke logging and eyetracking for investigating L2 writing fluency. *Studies in Second Language Acquisition, 41*(3), 583-604. https://doi.org/10.1017/S027226311900007X

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Duncheon, J. C., & Tierney, W. G. (2014). *Examining college writing readiness. In The Educational Forum* (Vol. 78, No. 3, pp. 210-230). Routledge. https://doi.org/10.1080/00131725.2014.912712

Feng, H. H., Saricaoglu, A., & Chukharev-Hudilainen, E. (2016). Automated Error Detection for Developing Grammar Proficiency of ESL Learners. *Calico Journal, 33*(1), 49-70. https://doi.org/10.1558/cj.v33i1.26507

Flower, L., & Hayes, J. (1980). The dynamics of composing: Making plans and juggling constraints. In L.W. Gregg & E.R. Steinberg (Eds.), *Cognitive Processes in Writing* (pp. 31-50). Hillsdale, NJ: Erlbaum.

Graham, S., Harris, K., & Hebert, M. A. (2011). *Informing writing: The benefits of formative assessment. A Carnegie Corporation Time to Act report.* L1 Research Archives Online.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*(1), 77-89. https://doi.org/10.1080/19312450709336664

Hayes, J. R., & Nash, J. G. (1996). On the nature of planning in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (p. 29–55). Lawrence Erlbaum Associates, Inc.

Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research, 1*(1). https://doi.org/10.17239/jowr-2008.01.01.1

Krippendorff, K. (2007). *Computing krippendorff's alpha reliability*. Departmental Papers (ASC), 43. https://doi.org/10.1002/9781405186407.wbiecr029

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication, 30*(3), 358-392. https://doi.org/10.1177/0741088313491692

Levelt, W. J. (1999). Producing spoken language: A blueprint of the speaker. In C.M. Brown and P. Hagoort (Eds.), *The neurocognition of language* (p. 83–122). https://doi.org/10.1093/acprof:oso/9780198507932.003.0004

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review, 8*(3), 299-325. https://doi.org/10.1007/BF01464076

National Center for Educational Statistics. (2011). *Writing 2011: National assessment of educational progress at grades 8 and 12.* Washington, DC: Institute of Educational Sciences, U.S. Department of Education.

National Commission on Writing. (2004). *Writing: A ticket to work…or a ticket out.* New York: The College Entrance Examination Board.

Olive, T. (2014). Toward a Parallel and Cascading Model of the Writing System: A Review of Research on Writing Processes Coordination. *Journal of Writing Research, 6*(2). https://doi.org/10.17239/jowr-2014.06.02.4

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics, 30*(4), 590-601. https://doi.org/10.1093/applin/amp045

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, 1*(3), 185-211. https://doi.org/10.1177/136216889700100302

Torrance, M., & Galbraith, D. (2006). The processing demands of writing. In C.A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 67–80). New York, NY: Guilford.

Vandermeulen, N., Leijten, M., & Van Waes, L. (2020). Reporting Writing Process Feedback in the Classroom Using Keystroke Logging Data to Reflect on Writing Processes. *Journal of Writing Research*, *12*(1), 109-139. https://doi.org/10.17239/jowr-2020.12.01.05

Van Galen, G. P. (1991). Handwriting: Issues for a psychomotor theory. *Human Movement Science, 10*(2-3), 165-191. https://doi.org/10.1016/0167-9457(91)90003-G

Van Waes, L., & Leijten, M. (2015). Fluency in writing: A multidimensional perspective on writing fluency applied to L1 and L2. *Computers and Composition, 38*, 79-95. https://doi.org/10.1016/j.compcom.2015.09.012

Van Waes, L., Leijten, M., Roeser, J., Olive, T., & Grabowski, J. (2021). Measuring and assessing typing skills in writing research. *Journal of Writing Research*, *13*(1), 107-153. https://doi.org/10.17239/jowr-2021.13.01.04

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

Appendix: Holistic Writing Rubric

Score	Quality of Work
5	Excellent and outstanding. The qualities of a B assignment, plus imagination, originality, and engaging expression. Writer responds thoughtfully and creatively, requiring very little or no revision. Both sides of the argument have been thoroughly addressed, and even the opposing viewpoint of the author has been discussed in terms of its arguments and counterarguments. Supporting details are relevant and provide important information about the topic.
4	Thorough analysis of and satisfactory solution to the communication task; good organization and solid expression. Writer responds fully, requiring some revision. Both sides of the argument have been addressed or recognized, though one side may still be more developed than the other.
3	Satisfactory analysis of the communication task, clear organization, and competent style; nothing remarkably good or bad. A 3 means the work meets the demands of the assignment in a minimally acceptable way. Writer responds mostly competently, requiring focused revision. Writer mainly addresses only one side of the argument; the other side is very briefly mentioned. Supporting details for the dominant argument are provided but may lack some depth.
2	Presence of a significant defect in context, substance, organization, style, or delivery in a lackluster paper; incomplete analysis of the communication task. Writer responds incompletely, requiring extensive revision. Writing may seem more like a haphazard collection of thoughts or notes than an argument or analysis of arguments. Some details are off-task; the prompt is only somewhat addressed.
1	Inadequate coverage of essential points, uncertain or misguided purpose, poor organization; ineffective and inconsistent expression; significant defects in standard usage; inadequate or inappropriate analysis of the communication task. Writer responds inadequately; paper is not acceptable.