



This is a repository copy of *The effect of keyboard-based word processing on students with different working memory capacity during the process of academic writing.*

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/116968/>

Version: Accepted Version

---

**Article:**

Van der Steen, S., Samuelson, D. and Thomson, J.M. [orcid.org/0000-0002-4293-4851](https://orcid.org/0000-0002-4293-4851)  
(2017) The effect of keyboard-based word processing on students with different working memory capacity during the process of academic writing. *Written Communication*, 34 (3). pp. 280-305. ISSN 0741-0883

<https://doi.org/10.1177/0741088317714232>

---

**Reuse**

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

**The effect of keyboard-based word processing on students with different working  
memory capacity during the process of academic writing**

Steffie Van der Steen

University of Groningen

Dianne Samuelson and Jennifer M. Thomson

Harvard Graduate School of Education

Author Note

Steffie Van Der Steen, Department Developmental Psychology, University of Groningen, Groningen, the Netherlands; Jennifer M. Thomson, Harvard Graduate School of Education, Harvard University; Dianne Samuelson, Harvard Graduate School of Education, Harvard University.

Steffie Van Der Steen is now at Department of Special Needs Education and Youth Care, University of Groningen, Groningen, the Netherlands. Jennifer M. Thomson is now at the Department of Human Communication Sciences, University of Sheffield, Sheffield, UK.

This research was supported in part by a grant from the Harvard University Milton Fund.

Correspondence concerning this article should be addressed to Jennifer M. Thomson, Department of Human Communication Sciences, University of Sheffield, Sheffield, S10 2TS, UK. Tel: +44 114 2222 440. Email: [j.m.thomson@sheffield.ac.uk](mailto:j.m.thomson@sheffield.ac.uk)

### **Abstract**

This study addresses the current debate about the beneficial effects of text processing software on students with different working memory (WM) during the process of academic writing, especially with regard to the ability to display higher-level conceptual thinking. 54 graduate students (M = 15, F = 39) wrote one essay by hand and one by keyboard. Our results show a beneficial effect of text processing software, in terms of both the qualitative and quantitative writing output. A hierarchical cluster analysis (HAC) was used to detect distinct performance groups in the sample. These performance groups mapped onto three differing working memory profiles. The groups with higher mean WM scores manifested superior writing complexity using a keyboard, in contrast to the cluster with the lowest mean WM. The results also point out that more revision during the writing process itself does not inevitably reduce the quality of the final output.

Keywords: Real-time Writing process; Text Processing Software; Manual Writing; Writing Quality; Skill Theory.

**The effect of keyboard-based word processing on students with different WM capacity during the process of academic writing**

Information technology is constantly transforming how knowledge is accessed and shared. The acquisition of knowledge is particularly critical within the field of education, where educational technology is increasingly used as a medium for both learning new information, as well as demonstrating knowledge, mostly—although not exclusively—in the form of written assignments using text processing software (Goldberg, Russell, & Cook, 2003). Especially for students struggling with a limited working memory (WM) capacity, technologies such as personal computers are generally viewed as advantageous. By reducing the mechanical demands of manual writing and facilitating revisions, capacity for higher-level processes is assumed to become available (Li & Hamel, 2003). However, Van Waes and Schellens (2003) point out that this same facility to revise at lower linguistic levels (e.g., correcting spelling or typing errors) might prevent writers from attending to revisions at higher levels, such as the structure of an argument (see also Haas, 1989). This means that keyboard-based word processing could provide an additional obstacle for students with a lower WM capacity. Although the number of typed (as opposed to written) assignments has increased in the last decades (Mogey, Cowan, Paterson, & Purcell, 2012), the relationship between WM and extended writing with or without technological assistance has never been explicitly investigated. The aim of this paper is therefore to fill this void, and empirically investigate how keyboard-based word processing affects students with different WM capacities during the process of academic writing.

### **Academic Writing and the Effect of Word Processing**

Academic writing is commonly used as a form of assessment in education, especially in postsecondary institutions, to let students demonstrate their critical thinking and ‘deep understanding’ of a certain topic (Mogey, et al., 2012). Writing is, however, one of the most complex skills demanded of students (Lindstrom, 2007). A possible reason for this is that writing involves multiple processes, including generating, organizing and synthesizing ideas, goal setting, the production of text, as well as the process of revision (Flower & Hayes, 1981; Hayes & Chenowith, 2006). To manage these processes, students need a variety of skills, such as planning abilities, retrieving information, generating text (knowledge of syntax, semantics), monitoring and revising text, etc. (Flower & Hayes, 1980; Torrance & Galbraith, 2006). Planning is particularly important; many thinking processes involved in writing are not automatized, and given the limited capacity of our cognitive system, this requires coordination (Torrance & Galbraith, 2006). Although the production of a single sentence, which involves word retrieval, developing a syntactic structure, retrieving phonology, and motor planning, can occur in an automatized fashion (Chanquoy, Foulin, & Fayol, 1990; Torrance & Galbraith, 2006), producing a full text requires more cognitive resources, as each sentence needs to be connected to both the previous and next sentence, with the overall structure of the argument kept in mind (Torrance & Galbraith, 2006).

In the last two decades, students’ academic writing is increasingly performed using electronic devices, such as a (laptop) computer and text processing software (Goldberg et al., 2003). Studies have found that, in general, word processing has medium-sized positive effects on writing quantity—meaning that students writing with a keyboard produce longer texts—and small- to medium-sized positive effects on writing quality (see Bangert-Drowns, 1993 and

Goldberg et al., 2003 for a meta-analysis). In Graham and Perin's (2007) meta-analysis on effective instructional practices to teach writing, word processing had a medium positive effect (Cohen's  $d = .55$ ) on students' writing quality, measured by an overall score that included factors such as the organization of the text, sentence structure, vocabulary, and tone. In addition, most college-age students can type faster than they can handwrite (MacKenzie & Soukoreff, 2002; Moge et al., 2012), and essay length is positively correlated with the quality of word-processed texts (Lovett, Lewandowski, Berger, & Gathje, 2010). Finally, using text processing software stimulates revision (Goldberg et al., 2003; Moge et al., 2012), and revisions made via keyboard potentially result in higher quality writing compared to revisions made using pen and paper (Goldberg et al., 2003). These positive findings suggest the potential effectiveness of word processing software for academic writing.

While research has shown that technology assists with lower-level writing processes such as proofreading, grammar, spelling and outlining (Li & Hamel, 2003), no empirical evidence explicitly suggests that these technologies facilitate higher-level conceptual writing skills needed to construct an argument. An early study by Haas (1989) showed that typing, as opposed to manual writing, obstructed the higher-level conceptual planning process of the participants (thoughts related to the structure and meaning of the text), while it facilitated sequential planning (focusing on lexical or syntactic characteristics). A possible reason for this might be that writers have more opportunities to change their text when using text processing software (cf. Case, 1985). Students seem to be aware of this as well. When Moge et al. (2012) asked students if they would prefer using word processing software for their written (in class) examinations, students responded that they worried that they would "waste time making lots of small but essentially trivial changes" (p. 120). In a recent study, Mueller and Oppenheimer (2014) found

that participants using a laptop while taking lecture notes performed significantly worse than participants who wrote manually on conceptual questions, but not on factual questions, suggesting that manual writing might enable people to process information at higher cognitive levels.

Only a handful of peer-reviewed studies to date have directly compared the real-time process of writing using a keyboard versus pen and paper (Berninger, Abbott, Augsburger & Garcia, 2009; Gould, 1981; Haas, 1989; Van Waes & Schellens, 2003). Van Waes and Schellens (2003) asked 40 experienced writers to write a technical report. Keystroke software collected process information from the keyboard condition while videotape capture was used to log the equivalent information from the pen and paper condition. Process data including pausing and revision behavior were analyzed. The use of a word processor, even by skilled post-graduate students, produced more fragmentary pause time patterns. The participants also tended to revise more extensively at the beginning of the writing process, and attended more to lower linguistic levels (letters, words). This suggests that the facility of lower level revisions provided by computers may in fact actively, “distract the writer’s attention from the possibility of revision at higher levels” (Van Waes & Schellens, 2003, p.833; see also Haas, 1989).

### **Working Memory Capacity and Academic Writing**

Regardless of the writing mode, the process of writing is assumed to be heavily mediated by the working memory (WM) system (Olive, 2014), which has led to an increased research interest in the last decades (Berninger & Winn, 2006; Bourke & Davies, 2013; Kellogg, 1996; McCutchen, 2000; Swanson & Berninger, 1996; Torrance & Galbraith, 2006; Vanderberg & Swanson, 2007). During writing, several cognitive processes access the limited resources of our WM system (Berninger & Winn, 2006; McCutchen, 2006; Olive, 2004; 2014; Swanson &

Berninger, 1996). WM is thought to be central to the full writing process, including processes of planning, idea and text generation, and revision (Hayes, 2000; McCutchen, 2000; Olive, 2004; 2014). This makes writing a demanding process, using much of the limited capacity of the WM, which sometimes causes an overload (Swanson & Berninger, 1996). It has been found, for example, that increases in cognitive load lead to more subject-verb agreement errors (Fayol, Largy, & Lemaire, 1994).

Throughout the literature on writing and cognitive resources, the conceptualization of WM largely depends on Baddeley's (1986) definition, stating that WM consists of two specialized memory systems, the visuospatial sketchpad (for spatial content) and phonological loop (for verbal content) to store information (Baddeley & Hitch, 1974). These are controlled by the central executive, that directs the flow of information in and out of these two specialized systems (Baddeley & Hitch, 1974; Cowan, 2016). Both phonological and visuospatial memory seem important for the writing process. While the role of phonological memory has been established in earlier studies (e.g., Berninger, 2009; Bourke & Adams, 2003), a recent study also found evidence for the role of the visuospatial working memory system in children's writing development (Bourke, Davies, Sumner, & Green, 2014). In the current study we use the work of Daneman and Carpenter (1980) and conceptualize WM not only as a temporary storage of phonological and visuospatial information, but also the processing of this information (see Cowan, 2016 for an interesting overview of WM conceptualizations). People with a high WM capacity are well-able to both retain information, as well as performing an additional operation on that information (Conway et al., 2005; Swanson, 1993; cf. Daneman & Carpenter, 1980), such as recalling the final words of each sentence after having heard a list of sentences (Daneman & Carpenter, 1983).



In order to become skilled, however, a writer also needs encoding methods and retrieval structures to quickly retrieve and use information stored in long-term memory. Ericsson and Kintsch (1995), based on their studies on expert performers, proposed that the working memory system includes another element, namely ‘long-term working memory’. This would allow for skilled use of accessible information storage in the long-term memory, and would hence extend WM capacity. More recently, Baddeley (2000) added an interim storage buffer to his WM model, the episodic buffer, which also integrates information from the WM systems with the long-term memory system (Berninger, Garcia, & Abbott, 2009). These revised models seem to be in line with research on the role of WM in writing (McCutchen, 2000). That is, while more advanced writers can move beyond the limited capacity of the WM system—possibly by making use of fluent encoding processes to retrieve knowledge from their long-term working memory—novice or struggling writers are hindered by the limited capacity of their WM system. For this latter category of students, reduction of cognitive task demands competing for limited WM capacity is thus an important goal.

Through this lens, text processing software is generally viewed as assistive. According to McCutchen (2006), manual writing may never be fully automatized, and hence, text processing is believed to reduce the mechanical demands, freeing up capacity for the higher level processes required for text generation (Berninger & Winn, 2006; Li & Hamel, 2003). However, as pointed out by Van Waes and Schellens (2003), the fact that word processors facilitate revisions on lower linguistic levels might actually prevent the writer from attending to revisions on higher levels, such as the structure of the argument, its content, or its complexity. Strikingly, using text processing might even obstruct the planning process before the first words are put on paper, even for experienced writers. Yet, this initial planning is considered important, as the writer starts to

organize the text and explores possible content and thesis (Haas, 1989). In a study on the writing process of children with learning disabilities, Berninger et al. (2009) report that for both children with and without learning disabilities, text processing software facilitated faster generation of separate letters, and sometimes even sentences, but not text generation, which puts higher demands on the WM system. Simply providing a computer, Berninger and colleagues conclude, is therefore not a solution for children struggling with writing assignments.

Thus, while keyboard-based word processing is generally viewed as assistive and increasing the writer's cognitive capacity for higher-level thinking, other researchers have suggested an opposite effect. This means that text processing software could provide an additional obstacle for students with low WM capacity, by cluttering their limited WM with more accessible revisions at the word- or sentence-level. The relationship between WM and writing with or without keyboard-based word processing has, however, never been empirically investigated in terms of both quantitative and qualitative writing output.

### **Measuring writing output**

Quantitative writing measures often cover the time spent writing, the number of words, a combination of these two (i.e. words per minute), or the number of grammatically correct sentences (Berninger et al., 2009). Special categories of quantitative writing measures are pause times and revision patterns (Van Waes & Schellens, 2003). Pauses and revisions tell us something about the real-time thinking process of the writer. More pausing, especially within a sentence, may signal a more fragmented and inefficient writing process (cf. Schilperoord, 1996).

The quality of writing is usually assessed with a holistic measure (Bangert-Drowns, 1993). This often consists of (independent) raters' assessment of the overall text, including its organization, sentence structure, the ideas conveyed, and the tone (Graham & Perin, 2007). In

the study of Troia and Graham (2002), for instance, writing quality was measured on a scale from 1 to 8 for both organization, and for the clarity and definition of the arguments. While there is no doubt that such measures of quality reflect the complexity of the written text to some extent, it remains questionable if they can fully assess higher cognitive complex thinking. Ideally, a measure of quality should also be equally applicable to a variety of texts and topics, making it possible to assess different texts on the same scale.

Skill Theory (Fischer, 1980; Fischer & Bidell, 2006) and its derivative the Lectical Assessment System, constructed to analyze test responses and essays (Dawson-Tunik, 2004), meet both criteria. Originally a theory of human cognitive development, Skill Theory proposes a hierarchical complexity scale along which people develop their understanding and other cognitive skills. The theory can be used to analyze people's performances in a wide variety of domains, by informing researchers on the construction of cognitive skill structures, which enables them to localize these performances along the complexity scale (Stein, Dawson, & Fischer, 2010). This complexity scale is hierarchically organized in three consecutive tiers that each contain three hierarchically connected levels. The three levels on the sensorimotor scale depict a concrete, basic understanding, which form the basis for the next, representational tier that specifies more complex relations between basic elements, which, in turn, serves as the foundation for the abstract tier, encompassing valid abstract rules (Fischer & Bidell, 2006; Van Der Steen, Steenbeek, Van Dijk, & Van Geert, 2014). Texts that show higher complexity levels generally connect more elements, specify relations between elements, and go beyond concrete instances by forming more abstract rules (Stein et al., 2010).

## **The Current Study**

Several researchers have called for more studies on the difference in cognitive demands between writing using a keyboard versus writing manually, especially when it comes to individual differences (Burke & Cizek, 2006; Chen, White, McCloskey, Soroui, & Chun 2011; Lindstrom, 2007). At the same time, and as outlined above, there exists a debate about the beneficial effects of text processing software on the WM demands during writing, especially with regard to the ability to display higher-level conceptual thinking. The goal of this study is therefore to empirically investigate how text processing software affects students with different WM capacities during the process of academic writing. This goal is addressed within the specific context of expository text, a complex and highly utilized genre within higher education (Mogey et al., 2012). We do so by focusing on quantitative patterns of writing behavior (measures of pauses and writing time) as well as qualitative measures reflecting the hierarchical complexity of the text (measures derived from the Lectical Assessment System —Skill Theory).

Three research questions specifically guided our study: 1) What is the general effect of modality (manual writing versus word processing software) on the qualitative and quantitative writing output? 2) Which distinct groups can be distinguished based on their quantitative and qualitative writing output, and what are the within-group effects of modality on the writing output? 3) How do these writing output groups differ in terms of their WM capacity?

## **Method**

### **Participants**

The sample was drawn from the general population of a U.S. graduate education program. All students in the graduate program received a recruitment email, which resulted in a

response rate of 19%. The recruited participants were subjected to three screening procedures. The first method of screening was an on-line survey which determined that all participants were native English speakers and used both writing modalities on a regular basis. Next, an individual interview was conducted to identify and exclude individuals diagnosed with health conditions that might affect the interpretability of the results, such as anxiety disorders triggered in a testing environment, epilepsy, or dysgraphia. Two individuals were excluded at this stage. Finally, participants undertook a behavioral assessment to screen for any significant graphomotor or literacy difficulties that would contra-indicate participation in the study. The Test of Word Reading Efficiency (TOWRE-2; Torgesen et al., 2011), Woodcock Johnson III Broad Written Language – Spelling subtest (Woodcock et al., 2001) and the Purdue Pegboard (Tiffin, 1968; 2002) for graphomotor dexterity were administered. Following the exclusion criteria of two standard deviations above or below the mean, no participants were excluded. The remaining 54 participants (M = 15, F = 39; M<sub>age</sub> 26, SD 2.79; age range: 21 to 34 years; 18 to 25 years of education) signed an informed consent. The study was approved by the local university ethics committee.

### **Memory Assessment Measures**

WM tasks are characterized by not only recalling presented information, but also performing an additional operation on that information (Conway et al., 2005; Swanson, 1993; cf. Daneman & Carpenter, 1980<sup>1</sup>). According to Swanson and Berninger (1996), who correlated several WM measures with writing skill, WM measures should include “language-related

---

<sup>1</sup> But see also the work of Cowan and colleagues (2005) suggesting that tasks without a combination of storage and processing are also correlated with aptitude measures.

processes” (p. 379) when assessing individual differences in writing performance. Other authors have mentioned the lack of a single “pure” working memory task (Conway et al., 2005). We therefore decided to administer two renowned tests of verbal WM and look at converging evidence from across these measures.

**WAIS-III Digit Span Backward.** This test (Wechsler, 2008) measured verbal working memory, specifically the phonological loop component (Baddeley & Hitch, 1974). A list of sequences with random numbers was read aloud to the participant. At the end of each sequence, the participant had to recall the numbers in the reversed order.

**Oral Sentence Span.** This test (Daneman & Carpenter, 1980, 1983) measured the central executive function of verbal working memory span. The examiner read aloud from the set a sentence every five seconds. The participant repeated the sentences aloud and was asked to recall the final word of each sentence from the set at the end, in any order.

## **Materials**

The participants were asked to write two essays responding to moral-dilemma writing prompts, one using a keyboard and one manually. The prompts were designed to be morally complex, and covered themes that require equal background knowledge. Participants were recruited from a graduate school of education, and so one prompt focused upon an inappropriate teacher-student relationship, while the other focused upon a case of students changing their grades as a result of computer hacking. In both cases, the participant was asked to indicate how strongly they agreed or disagreed with the deliberately ambiguous outcomes described, and asked to provide a written rationale for their response. Every effort was made to equate the conceptual and written demands of each prompt and so both were of equivalent length, with similarly structured response requirements. Alongside this structural equivalence, both prompts

were analyzed by two researchers with Lectical Assessment expertise who were blind to the purposes of the study; the researchers were used to confirm that responses to each prompt could be considered from a comparable number of perspectives, i.e. could contain similar levels of complexity.

### **Procedure**

Each participant attended two experimental sessions. In each session, the participant was given a prompt to read. The participants were randomly assigned to both a writing modality and prompt for the first test session, and administered the remaining prompt in the other modality for the second session (prompts were randomized across the two modalities). The participants were advised to use 45 minutes to answer, but if additional time were required, it would be given. No one exceeded one hour. To prevent fatigue, all participants had several days at a minimum between tests.

For the experimental condition of writing manually participants used the Intuos 4 Wacom tablet, with a lined self-adhesive paper placed on top of the tablet workspace and an inking pen stylus. By using Eye and Pen software (Alamargot et al., 2006; Chesnet & Alamargot, 2008), every pen stroke was recorded and time stamped to within a millisecond. Eye and Pen also measured the pressure level of the pen permitting the capture of up/down states, providing pause durations.

In the keyboard condition, a Dell laptop computer loaded with the software program ScriptLog (Stromqvist & Karlsson, 2002) was used. On the screen appeared a text box similar in appearance to most word processing software. The menu of features included the ability to delete, move throughout the document with the mouse or cursor, a return, and capitalization. Spell and grammar checks were not available. ScriptLog recorded every keystroke and keyboard

event to the millisecond. In both conditions, participants performed small tasks to get used to the software (type/write their name and the alphabet), before beginning the experimental session.

### **Data Analysis**

**Writing output.** Both ScriptLog and Eye and Pen enabled the researchers to replay the writing process and provided the following quantitative data: total writing time (measured from the first pen- or key-stroke), total number of words, words generated per minute, and the ratio of pause time (pauses longer than 2 seconds) to active writing time. In addition, the essays were subject to an analysis of the complexity of their argumentation using the Lectical Assessment System (Dawson, 2010), which is a psychometrically refined metric based on Skill Theory (Fischer, 1980; Fischer & Bidell, 2006). Two expert LAS analysts independently scored each essay as a whole, in order to yield a total complexity score (Dawson & Stein, 2010). The analysts received typed copies of the responses from both conditions, in order to keep them blind as to the modality of the response they were reading.

The Lectical Assessment System (LAS) assesses students' performance by examining the level of complexity of the concepts present in the argumentation, and the logical structure of the argument (Dawson-Tunik 2004; 2006; Dawson & Wilson, 2004). The LAS metric is particularly valuable in the cases of open-ended questions and dilemmas (Stein & Heikkinen, 2009), and is considered domain-general. Hence, the scoring is not bound to a particular task or domain (Dawson & Wilson, 2004). The LAS metric has exhibited high inter-rater reliability (80-97% agreement). With regard to the validity, the LAS corresponds well with scores assigned by other developmental metrics, such as Kohlberg's Standard Issue Scoring System for morality; Armon's Good Life Scoring System for evaluative reasoning; Perry's Scoring System for epistemological understanding, and Kitchener and King's Reflective Judgment Scoring System



(agreement rates of 85% or greater; Dawson, Xie & Wilson, 2003; Stein & Heikkinen, 2009).

For the current project, the initial rate of agreement between two raters within 1/4 of a level was 85.7%. After discussion of differences larger than 1/3 of a level, it was 95.5%. All essays were then scored by these two raters. We took their average score for further analysis.

**General effect of modality on writing output (research question 1).** We performed a repeated measures Anova to examine the within-subject effect of modality on the writing output measures. A power analysis in G Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that 39 participants should be included for this analysis to have 95% power for detecting a small sized effect, adhering to the .05 criterion of statistical significance. The observed p-value was corrected by using the Bonferroni correction (multiplied by the number of comparisons). For each writing output measure, an effect size in the form of the partial  $\eta^2$  was calculated.

**Which groups can be distinguished based on writing output? (research question 2).** To see if this multivariate dataset comprised distinct groups with different performance patterns, we used Hierarchical Agglomerative Clustering (HAC) in the statistical program Tanagra 1.4.41 (Rakotomalala, 2005). We used participants' standardized LAS complexity scores, total writing time, ratio of pause/writing, total number of words, and words per minute for each modality as the input variables of the clustering. The HAC method successively merges individuals who show a similar pattern of scores into groups (i.e. clusters), keeping the within-cluster variance at a minimum. The clustering starts with all individual subjects, joining the two most similar subjects together, after which the next two most similar are merged. From the resulting hierarchy, the appropriate number of clusters needs to be determined. For this, the gap statistic (Tibshirani et al., 2002) is widely used. The gap statistic compares the change in within-cluster dispersion with what would be expected under a null distribution, e.g., no clusters (Tibshirani et

al., 2002; Martinez, Martinez, & Solka, 2004). We proceeded by validating the solution of the HAC using a K-means clustering technique (Hartigan & Wong, 1979; Steinley, 2006), to see if it provided the same number of clusters, and by comparing the two solutions with the Chi-square statistic.

We then used the group characterization function in Tanagra 1.4.41 (Rakotomalala, 2005), to see if the clustering resulted in groups with lower and higher writing output scores. This group characterization was formalized in test values that show how much weight each variable has in a specific cluster, based on a comparison of the mean of each variable within a specific cluster (e.g., the mean writing time within cluster 1) with the mean of this variable across the whole dataset (e.g., the mean writing time within the whole sample). The test value asymptotically follows a Gaussian distribution, and absolute values greater than 2 signal that the value of the variable in a specific cluster is significantly different from its value in the rest of the sample (Rakotomalala, 2005).

Meaningful differences between the clusters, signaled by their test values, were further explored using Monte Carlo permutation tests to determine the p-values. This method is particularly useful in the case of small sample sizes or unbalanced data sets (Anderson, 2013; Todman & Dugard, 2001). Taking the sample distribution into account, a Monte Carlo test measures the probability that a difference is caused by chance. This is done by drawing (in this case) 5000 random samples from the original data, and determining how often the observed or a bigger difference occurs in these random samples. This number is then divided by the number of drawn samples (5000), which produces a p-value.

### **Characterization of groups in terms of working memory (research question 3).**

Lastly, we used Monte Carlo permutation tests to examine the association between the clustering

(based on the writing output measures), and the three working memory measures from the screening. Using this method, it was possible to see if observed trends in the clustering based on the writing output (i.e. clusters with higher and lower scores) could be linked to specific trends in the WM measures.

For both research question 2 and 3, where separate groups were compared, we calculated the effect size (Cohen's  $d$ ) by dividing the observed difference between groups by the pooled SD. A value of  $d$  between 0.2 and 0.3 is generally considered as a small effect, a value around 0.5 as medium, and a value of 0.8 and higher is considered as large (Cohen, 1988).

## Results

### General Effect of Modality on Writing Output (Research Question 1)

Table 1 shows the summary statistics for written output variables of the repeated measures Anova. In general, all students received a higher complexity score in the keyboard condition ( $F(1,53) = 8.39, p = .025, \eta^2 = .14$ ); they also wrote significantly more words ( $F(1,53) = 31.59, p < .001, \eta^2 = .37$ ), and generated more words per minute ( $F(1,53) = 38.38, p < .001, \eta^2 = .42$ ). The ratio of pause/writing was, however, significantly higher in the keyboard condition, meaning that students spent more time pausing instead of actively generating text ( $F(1,53) = 24.29, p < .001, \eta^2 = .31$ ). Although the total writing time also seemed to be longer in the keyboard condition, no significant difference between the modalities was found. In sum, the results seem to point to a slight advantage for the keyboard condition (more words, faster writing, and a higher complexity score), apart from the ratio pause/writing.

### Which Groups Can Be Distinguished Based on Writing Output? (Research Question 2)

Using hierarchical agglomerative clustering (HAC), three groups with a distinct writing output could be identified ( $n_1 = 25$ ,  $n_2 = 14$ , and  $n_3 = 15$ ;  $R^2 = .45$ ), with a gap statistic of 1.03. To check the validity of the clustering, a K-means clustering was performed, which resulted in an almost exact copy of 3 clusters ( $n_1 = 24$ ,  $n_2 = 14$ , and  $n_3 = 16$ ;  $\chi^2_{\text{K-means-HAC}}(2, N = 54) = 102.6$ ;  $p < .001$ ).

Cluster 1 scores significantly higher on most variables for both modalities, apart from the ratio of pause/writing in the manual writing condition ( $M = .31$ ), on which this cluster scores below average (test value = -1.89,  $d = .35$ ), meaning that they spend less time pausing in this modality. Table 2, which contains the test values per variable, shows that this cluster often scores higher than the other two clusters, as indicated by the high positive test values (for the means and standard deviations, see Table 3). The participants in cluster 2 have a significantly higher ratio of pause/writing time (i.e. they spend more time pausing) in both modalities ( $M_{\text{pen}} = .49$ , test value = 5.73,  $d = 3.35$ ;  $M_{\text{keyboard}} = .5$ , test value = 2.4,  $d = .88$ ). They spend significantly more time writing in the keyboard condition than the other clusters ( $M = 28.91$ , test value = 1.2,  $d = .43$ ), but not in the manual writing condition ( $M = 26.09$ , test value = .58), and generate less text per minute in both modalities ( $M_{\text{pen}} = 13.52$ , test value = -4.76,  $d = 2.27$ ;  $M_{\text{keyboard}} = 16.61$ , test value = -3.89,  $d = 1.63$ ). Their complexity scores are not significantly different from the other two clusters ( $M_{\text{pen}} = 11.08$ , test value = -1.64;  $M_{\text{keyboard}} = 11.26$ , test value = -.04). This cluster is thus characterized by both positive and negative test values, less extreme than the other two clusters. Cluster 3 generates significantly more words per minute than the other clusters for both modalities ( $M_{\text{pen}} = 21.25$ , test value = 2.97,  $d = .99$ ;  $M_{\text{keyboard}} = 27.68$ , test value = 2.4,  $d = .77$ ), and spends less time pausing when writing manually ( $M = .25$ , test value = -3.49,  $d = 1.22$ ). The complexity scores of participants in cluster 3 are, however, significantly lower than the other two

clusters ( $M_{\text{pen}} = 11.0$ , test value = -3.38,  $d = 1.16$ ;  $M_{\text{keyboard}} = 11.04$ , test value = -3.87,  $d = 1.39$ ), as well as their total number of words ( $M_{\text{pen}} = 273.6$ , test value = -4.22,  $d = 1.59$ ;  $M_{\text{keyboard}} = 415.5$ , test value = -3.14,  $d = 1.06$ ), and the writing time in both modalities ( $M_{\text{pen}} = 13.25$ , test value = -5.64,  $d = 2.76$ ;  $M_{\text{keyboard}} = 16.16$ , test value = -4.51,  $d = 1.76$ ). In fact, Table 2 shows that the test values of this cluster are often negative, indicating that they score lower on the writing output measures than the rest of the sample.

Table 3 takes a closer look at the within-cluster modality differences, and shows that the slight advantage for the keyboard condition that was found across the sample, is also visible within each cluster. Cluster 1 has a higher complexity score ( $M_{\text{pen}} = 11.3$ ;  $M_{\text{keyboard}} = 11.4$ ;  $p = .02$ ,  $d = .59$ ), higher number of words ( $M_{\text{pen}} = 595.64$ ;  $M_{\text{keyboard}} = 745.24$ ;  $p < .01$ ,  $d = .85$ ), and generated more words per minute in the keyboard condition ( $M_{\text{pen}} = 19.35$ ;  $M_{\text{keyboard}} = 25.01$ ;  $p < .01$ ,  $d = 1.14$ ). The total writing time is, however, roughly equal in both conditions ( $M_{\text{pen}} = 31.09$ ;  $M_{\text{keyboard}} = 30.56$ ;  $p = .63$ ), and the ratio pause/writing is higher in the keyboard condition ( $M_{\text{pen}} = .31$ ;  $M_{\text{keyboard}} = .42$ ;  $p < .01$ ,  $d = 1.16$ ), indicating participants spent more time pausing while writing using word processing software. Cluster 2 shows roughly the same pattern, apart from the ratio of pause/writing time, which is not significantly different for the two modalities ( $M_{\text{pen}} = .49$ ;  $M_{\text{keyboard}} = .5$ ;  $p = .38$ ). Cluster 3 does not show a significant difference in LAS complexity scores between the two modalities ( $M_{\text{pen}} = 11.0$ ;  $M_{\text{keyboard}} = 11.04$ ;  $p = .31$ ). These participants seem to write longer in the keyboard condition ( $M_{\text{pen}} = 13.25$ ;  $M_{\text{keyboard}} = 16.16$ ;  $p = .09$ ), although this is not statistically significant. Just like cluster 1, this cluster spent more time pausing in the keyboard condition ( $M_{\text{pen}} = .25$ ;  $M_{\text{keyboard}} = .4$ ;  $p < .01$ ,  $d = 1.57$ ).

Generally speaking, cluster 1 seems to perform well in both modalities compared to the other clusters, cluster 3 performs worse and cluster 2 falls in between. An advantage for the

keyboard condition is found for all clusters, although none shows a difference in total writing time between the two modalities, and the time spent pausing is generally higher in the keyboard condition. With regard to writing complexity, only clusters 1 and 2 manifest a significant effect of modality, i.e. superior writing complexity using a keyboard.

### **Characterization of Groups in terms of Working Memory (Research Question 3)**

In Table 4 the clusters are linked to the WM measures obtained during the screening. We tested whether the patterns found in the clustering of the writing output measure—one high-performing, one average, and one low-performing cluster—were reflected in the patterns of their WM measures. Cluster 1 scored higher than cluster 2 on the digit backward span ( $M_{diff} = .63$ ), which in turn, scored higher than cluster 3 ( $M_{diff} = .42$ ). A Monte Carlo analysis revealed that these differences between the groups were significant ( $p_{cluster1>2>3} = .02$ ,  $d = .65$ ). In addition, cluster 1 had higher values on the oral sentence span measure than cluster 2 ( $M_{diff} = 1.91$ ), and cluster 2 scored slightly higher than cluster 3 on this measure ( $M_{diff} = .1$ ). A Monte Carlo analysis revealed that these differences between the groups were significant ( $p_{cluster1>2>3} < .01$ ,  $d = 1.94$ ).

In sum, the patterns found in the clustering of the writing output measures were reflected in the patterns of digit backward span and oral sentence span. The results show that people who perform well on the writing tasks also score high on the WM tasks (cluster 1). Moreover, the group that scores particularly low on the writing indicators, also scores low on the WM measures (cluster 3).

## **Discussion**

This study set out to investigate how text-processing software affects students with different WM capacities during the process of academic writing. The goal was addressed within

the specific context of complex, expository text. Across the participants as a whole, significant within-subject differences were found in both quantitative patterns of writing behavior as well as qualitative measures reflecting the complexity of the text: Students wrote more words, and at a faster rate in the keyboard condition, as compared to the pen and paper condition, and in terms of final output, the essays written on the keyboard were of significantly greater conceptual complexity. These overall findings are in accord with previous studies: The two key meta-analyses of Bangert-Drowns (1993) and Goldberg et al. (2003) both report greater output and higher quality of output for keyboard writing as opposed to that generated by pen and paper. Equally, a faster rate of typing than hand-writing has previously been documented for college-age students (Mogey et al., 2012).

The current study went one step further by examining whether this global profile is influenced by individual differences in working memory, a factor known to play a significant role in effective writing (McCutchen, 2000). Using hierarchical agglomerative clustering (HAC), three distinct performance groupings were identified that mapped onto three differing working memory profiles. Replicating the findings of previous studies (see McCutchen, 1996 for a review), working memory performance was related to writing quality – the cluster (cluster 1) with higher means on a range of working memory measures also exhibited higher means in their LAS complexity scores, while the cluster with the lowest means on working memory tasks (cluster 3) achieved lower means in their complexity scores.

More uniquely, this study allowed us to see whether the higher writing quality scores in the keyboard condition occurred uniformly across the different clusters. This quality (LAS complexity) advantage for essays written by keyboard was observable for clusters 1 and 2, but not for cluster 3, which was the group with the lower working memory means. This cluster also

performed below the overall mean on most writing output measures for both typing and writing. Given that neither of the modalities resulted in an advantage for these students, we can deduce that the writing process itself might place a burden on these students' WM capacity (cf. McCutchen, 2000; 2006). For these students it would be helpful to investigate which aspects of the writing process are most affected by their WM capacity, which could help in constructing interventions aimed at the reduction of cognitive task demands competing for limited WM capacity while writing. As Swanson and Berninger (1996) put it, expert writers are less hindered by the limited capacity of their WM, as they have automatized many elementary writing processes, such as the structure of discourses and lexical access, leaving more capacity for generating and organizing ideas. It has been suggested that more advanced writers can move beyond the limited capacity of the WM system, possibly by making use of fluent encoding processes to retrieve knowledge from their long-term working memory (Ericsson & Kintsch, 1995; McCutchen, 2000; 2006). However, given that the students in cluster 3 had a similar educational background compared to the other two clusters, it is questionable whether their performance is a matter of experience with writing, and that their limited WM capacity could be overcome by simply building expertise.

Regarding the total amount of text written, and writing rate, the three clusters all mirrored the main effects of more text and faster writing in the keyboard condition. The keyboard condition also elicited a higher pause/writing time ratio (i.e. more time spent pausing) in the overall group and in clusters 1 and 3, while cluster 2 spent more time pausing in both modalities in general. According to the literature, writing on a keyboard would appear to place different demands on the WM system as compared to writing using a pen (Berninger & Winn, 2006; Berninger et al., 2009). As opposed to pre-planning and then writing, the use of computers



appears to encourage less initial planning and more thinking and revision once writing has begun (Haas, 1989; Van Waes & Schellens, 2003), which would be reflected in longer pause times. Indeed, the use of text processing software, with its user-friendly options to delete and retype text, seems to afford (cf. Gibson, 1977) revisions during the act of writing at the word- or sentence-level (Case, 1985; Mogyey et al., 2012). In the current study, an increase in the time spent pausing for the keyboard condition was also seen, but equally, an overall quality advantage for essays written by keyboard was observable across the group. Thus it appears that more pausing during the writing process itself does not inevitably reduce the quality of the final output.

### **Strengths, limitations and future directions**

The HAC uses a bottom-up procedure to cluster participants by using patterns in the multivariate data as the only input. We consider this an asset, as groups did not have to be determined a priori, and could be deduced based on patterns in no less than 10 writing output variables. While the HAC procedure yielded three groups with differing WM profiles, including a group with lower WM scores compared to the rest of the sample, note that our sample came from an institution for higher education, and was thus a high-performing group with considerable experience in writing expository texts. It would therefore be interesting to extend this study to other population groups, including less experienced groups, as well as students diagnosed with WM deficits, to see if the advantages of the keyboard condition over the pen and paper condition would still hold.

The program Scriptlog (Stromqvist & Karlsson, 2002) was used to record writing behavior in the keyboard condition. Despite that it is similar in appearance to most word processing software, Scriptlog has limited text editing functions, and although we gave

participants time to familiarize themselves with the program (see Methods section), they might have needed time to adapt. New keystroke logging programs with more editing options are currently available to overcome this (Leijten & van Waes, 2013).

Earlier research suggests that writing using a keyboard would elicit revision, mainly at lower linguistic levels (e.g., correcting spelling or typing errors). This would lead to an increase in pauses, but might also prevent writers from attending to revisions at higher levels (Van Waes & Schellens, 2003). In the current study, an increase in the time spent pausing for the keyboard condition was also seen, but equally, an overall quality advantage for essays written by keyboard. An important next step in this work will be studying the relationship between the specific location of pauses in writing (within a word, within a sentence, at sentence or at paragraph boundaries) and the characteristics and complexity of the final written output. More pauses within a word or sentence may signal a more fragmented and inefficient writing process (Schilperoord, 1996; Van Waes & Schellens, 2003). While doing so, one must take into account that pausing while using pen and paper may be different from pausing during keyboard text processing (cf. Van Waes, Leijten, Wengelin, & Lindgren, 2012). Moreover, pauses may not only reflect anticipatory planning of the next word(s), but may also represent the delayed effects of the previous word written, at least while writing manually (Maggio, Lété, Chenu, Jisa, & Fayol, 2012).

In future research, it will also be important to consider the relationship between WM and writing processes across different types of writing activity. In contemporary models of WM and writing relationships, WM is conceptualized as a mediator of writing process coordination (Kellogg, 1996; Olive, 2014). When the accumulated concurrent demands of writing processes exceed the limited WM capacity, then performance is compromised. However, this leaves many

unanswered questions in terms of the mechanism of this performance compromise. Are specific writing processes more vulnerable to WM overload, exerting a cascading effect upon subsequent processes, or do WM capacity issues have direct parallel effects across multiple writing processes concurrently? Given the ubiquity of writing within digital environments and the shift in writing processes this is precipitating, as seen in this study, we strongly recommend that future theory-building regarding relationships between working memory and writing is alert to writing modality and inclusive of digital modes.

### References

- Alamargot, D., Chesnet, D., Dansac, C., & Ros, C. (2006). Eye and pen: A new device for studying reading during writing. *Behavioral Research Methods*, 38, 287-299. doi: 10.3758/BF03192780
- Anderson, R. B. (2013). Conceptual distinction between the critical p value and the type I error rate in permutation testing. *Journal of Modern Applied Statistical Methods*, 12, 2-8.
- Baddeley, A.D. (1986). Working memory and comprehension. In D. Broadbent, J. McGaugh, M. Kosslyn, N. Mackintosh, E. Tulving, & L. Weiskrantz (Eds.), *Working memory* (pp. 75-108). Oxford, U.K.: Oxford University Press.
- Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences*, 4, 417-423. doi: 10.1016/S1364-6613(00)01538-2
- Baddeley, A.D. and Hitch, G.J. (1974) Working memory. In G. A. Bower. (Ed.), *The Psychology of Learning and Motivation* (pp. 47-89). Waltham, MA: Academic Press.
- Bangert-Drowns, R. L. (1993). The word processor as an instructional tool: A meta-analysis of word processing in writing instruction. *Review of Educational research*, 63, 69-93. doi: 10.3102/00346543063001069

Berninger, V. W. (2009). Highlights of programmatic, interdisciplinary research on writing.

*Learning disabilities: Research and Practice*, 24, 68–79. doi: 10.1111/j.1540-

5826.2009.00281.x

Berninger, V. W., Abbott, R. D., Augsburger, A., & Garcia, N. (2009). Comparison of pen and

keyboard transcription modes in children with and without learning disabilities. *Learning*

*Disability Quarterly*, 32, 123-141. doi: 10.2307/27740364

Berninger, V., Abbott, R., Jones, J., Wolf, B., Gould, L., Anderson-Youngstrom, M. et al.

(2006). Early development of language by hand: Composing-, reading-, listening-, and speaking-connections, three letter writing modes, and fast mapping in spelling.

*Developmental Neuropsychology*, 29, 61-92. doi: 10.1207/s15326942dn2901\_5

Berninger, V. W., Garcia, N. P., & Abbott, R. D. (2009). Multiple processes that matter in

writing instruction and assessment. In G. A. Troia (Ed.), *Instruction and assessment for struggling writers: Evidence-based practices* (pp. 15-74). New York: Guilford Press.

Berninger, V. W., & Winn, W. D. (2006). Implications of advancements in brain research and

technology for writing development, writing instruction, and educational evolution. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 96-114).

New York: Guilford Press.

Bourke, L., & Adams, A. M. (2003). The relationship between working memory and early

writing assessed at the word, sentence and text level. *Educational and Child Psychology*, 20, 19-36.

Bourke, L., Davies, S. J., Sumner, E., & Green, C. (2014). Individual differences in the

development of early writing skills: testing the unique contribution of visuo-spatial working memory. *Reading and Writing*, 27, 315-335. doi: 10.1007/s11145-013-9446-3

- Burke, J. N., & Cizek, G. J. (2006). Effects of composition mode and self-perceived computer skills on essay scores of sixth graders. *Assessing Writing*, 11, 148-166. doi: 10.1016/j.asw.2006.11.003
- Case, D. (1985). Processing professorial words: Personal computers and the writing habits of university professors. *College Composition and Communication*, 36, 317-322. doi: 10.2307/357975
- Chanquoy, L., Foulin, J. N., & Fayol, M. (1990). Temporal management of short text writing by children and adults. *Current Psychology of Cognition*, 10, 513-540.
- Chen, J., White, S., McCloskey, M., Soroui, J., & Chun, Y. (2011). Effects of computer versus paper administration of an adult functional writing assessment. *Assessing Writing*, 16, 49-71. doi: 10.1016/j.asw.2010.11.001
- Chesnet, D., & Alamargot, D. (2008). *Eye and Pen: User manual, version 1.1*. Octares Edition, National Center for Scientific Research, University of Poitiers.
- Christensen, C. A. (2004). Relationship between orthographic-motor integration and computer use for the production of creative and well-structured written text. *British Journal of Educational Psychology*, 74, 551-565. doi: 10.1348/0007099042376373
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cowan, N. (2016). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, 1-13. doi: 10.3758/s13423-016-1191-6
- Cowan, N., Elliott, E.M., Saults, J.S., Morey, C.C., Mattox, S., Hismjatullina, A., & Conway, A.R.A. (2005). On the capacity of attention: Its estimation and its role in working memory

and cognitive aptitudes. *Cognitive Psychology*, 51, 42-100. doi:

10.1016/j.cogpsych.2004.12.001

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W.

(2005). Working memory span tasks: A methodological review and user's guide.

*Psychonomic bulletin & review*, 12, 769-786. doi: 10.3758/BF03196772

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and

reading. *Journal of verbal learning and verbal behavior*, 19, 450-466. doi:10.1016/S0022-

5371(80)90312-6

Daneman, M. & Carpenter, P.A. (1983). Individual differences in integrating information

between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and*

*Cognition*, 9, 561-584. doi: 10.1037/0278-7393.9.4.561

Dawson-Tunik, T. L. (2004). "A good education is..." The development of evaluative thought

across the life-span. *Genetic, Social, and General Psychology Monographs*, 130, 4-112. doi:

10.3200/MONO.130.1.4-112

Dawson-Tunik, T. L. (2006). The meaning and measurement of conceptual development in

adulthood. In C. Hoare (Ed.), *The intersection of adult development and learning: A*

*handbook of theory, research, and practice* (pp. 433-454). London: Oxford.

Dawson, T.L., & Stein, Z. (2010). Developmental complexity and writing modality. A report for

J.M. Thomson, September 28, 2010.

Dawson, T. L., & Wilson, M. (2004). The LAAS: A computerized developmental scoring system

for small- and large-scale assessments. *Educational Assessment*, 9, 153-191. doi:

10.1080/10627197.2004.9652963

- Dawson, T. L., Xie, Y., & Wilson, M. (2003). Domain-general and domain-specific developmental assessments: do they measure the same thing?. *Cognitive Development*, 18, 61-78. doi: 10.1016/S0885-2014(02)00162-4
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological review*, 102, 211-245. doi: 10.1037/0033-295X.102.2.211
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G\* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39, 175-191. doi: 10.3758/BF03193146
- Fayol, M., Largy, P., & Lemaire, P. (1994). Cognitive overload and orthographic errors: When cognitive overload enhances subject-verb agreement errors. A study in French written language. *The Quarterly Journal of Experimental Psychology*, 47, 437-464. doi: 10.1080/14640749408401119
- Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological review*, 87, 477-531. doi: 10.1037/0033-295X.87.6.477
- Fischer, K. W., & Bidell, T. R. (2006). Dynamic development of action, thought, and emotion. In R. M. Lerner (Ed.) & W. Damon (Series ed.), *Handbook of Child Psychology: Vol. I. Theoretical Models of Human Development* (pp. 313-399). New York: Wiley.
- Flower, L., & Hayes, J. R. (1980). The cognition of discovery: Defining a rhetorical problem. *College composition and communication*, 31, 21-32. doi: 10.2307/356630
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College composition and communication*, 32, 365-387. doi: 10.2307/356600
- Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67-82). Hillsdale, NJ: Erlbaum

- Goldberg, A., Russell, M., & Cook, A. (2003). The effect of computers on student writing: A meta-analysis of studies from 1992 to 2002. *The Journal of Technology, Learning and Assessment*, 2, 1-52.
- Gould, J. D. (1981). Composing letters with computer-based text editors. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 23, 593-606. doi: 10.1177/001872088102300509
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of educational psychology*, 99, 445-476. doi: 10.1037/0022-0663.99.3.445
- Haas, C. (1989). How the writing medium shapes the writing process: Effects of word processing on planning. *Research in the Teaching of English*, 23, 181-207.
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28, 100-108.
- Hayes, J. R. (2000). A New Framework for Understanding Cognition and Affect in Writing. In R. Indrisano & J. R. Squire (Eds.), *Perspectives on writing: Research, theory, and practice* (pp.6-44). The International Reading Association. London, UK: Routledge.
- Hayes, J. R., & Chenoweth, N. A. (2006). Is working memory involved in the transcribing and editing of texts? *Written Communication*, 23, 135-149. doi: 10.1177/0741088306286283
- Horkay, N., Bennett, R. E., Allen, N., Kaplan, B., & Yan, F. (2006). Does It Matter if I Take My Writing Test on Computer? An Empirical Study of Mode Effects in NAEP. *Journal of Technology, Learning, and Assessment*, 5. Retrieved from <http://www.jtla.org>
- Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 57-71). Hillsdale, UK: Lawrence Erlbaum Associates



- Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication*, 30, 358-392. doi: 10.1177/0741088313491692
- Li, H., & Hamel, C. M. (2003). Writing issues in college students with learning disabilities: A synthesis of the literature from 1990 to 2000. *Learning Disability Quarterly*, 26, 29-46. doi: 10.2307/1593683
- Lindstrom, J. H. (2007). Determining appropriate accommodations for postsecondary students with reading and written expression disorders. *Learning Disabilities Research & Practice*, 22, 229-236. doi: 10.1111/j.1540-5826.2007.00251.x
- Lovett, B. J., Lewandowski, L. J., Berger, C., & Gathje, R. A. (2010). Effects of Response Mode and Time Allotment on College Students' Writing. *Journal of College Reading and Learning*, 40, 64-79. doi: 10.1080/10790195.2010.10850331
- McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational psychologist*, 35, 13-23. doi: 10.1207/S15326985EP3501\_3
- McCutchen, D. (2006). Cognitive factors in the development of children's writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 115-130). New York: Guilford Press.
- MacKenzie, I. S., & Soukoreff, R. W. (2002). Text entry for mobile computing: Models and methods, theory and practice. *Human-Computer Interaction*, 17, 147-198. doi: 10.1080/07370024.2002.9667313
- Maggio, S., Lété, B., Chenu, F., Jisa, H., & Fayol, M. (2012). Tracking the mind during writing: Immediacy, Delayed, and Anticipatory Effects on Pauses and Writing Rate. *Reading and Writing*, 25, 2131-2151. doi: 10.1007/s11145-011-9348-1

- Mangen, A., Anda, L., Oxenburgh, G., & Brønning, K. (2015). Handwriting versus keyboard writing: effect on word recall. *Journal of Writing Research*, 7, 227-247. doi: 10.17239/jowr-2015.07.02.1
- Martinez, W. L., Martinez, A., & Solka, J. (2004). *Exploratory data analysis with MATLAB*. Boca Raton, FL: CRC Press.
- Mogey, N., Cowan, J., Paterson, J., & Purcell, M. (2012). Students' choices between typing and handwriting in examinations. *Active Learning in Higher Education*, 13, 117-128. doi: 10.1177/1469787412441297
- Mueller, P. A., & Oppenheimer, D. M. (2014). The Pen Is Mightier Than the Keyboard Advantages of Longhand Over Laptop Note Taking. *Psychological Science*. doi: 10.1177/0956797614524581.
- Olive, T. (2004). Working memory in writing: Empirical evidence from the dual-task technique. *European Psychologist*, 9, 32-42. doi: 10.1027/1016-9040.9.1.32
- 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, 173-194. doi:10.17239/jowr-2014.06.02.4
- Rakotomalala, R., 2005. TANAGRA: a free software for research and academic purposes. In *Proceedings of EGC'2005, RNTI-E-3*, vol. 2. (pp 697-702).
- Schilperoord, J. (1996). The distribution of pause time in written text production. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Theories, models and methodology in writing research* (pp. 21–35). Amsterdam: Amsterdam University Press.
- Sheslow, D., & Adams, W. (2003). *Wide Range Assessment of Memory and Learning—Revised (WRAML-2)*. Administration and Technical Manual. Wilmington, DE: Wide Range. Inc.:

- Stein, Z., Dawson, T., & Fischer, K. W. (2010). Redesigning testing: Operationalizing the new science of learning: Cognition, Computers and Collaboration in Education. In M. S. Khine & I. M. Saleh (Eds.), *New Science of Learning* (pp. 207-224). New York: Springer. doi: 10.1007/978-1-4419-5716-0\_10
- Stein, Z., & Heikkinen, K. (2009). Models, metrics, and measurement in developmental psychology. *Integral Review*, 5, 4-24.
- Steinley, D. (2006). K-means clustering: a half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, 59, 1-34. doi: 10.1348/000711005X48266
- Strömquist, S., & Karlsson, H. (2002). *ScriptLog for windows: Users manual*. Lund, Sweden: University of Lund, Department of Linguistics.
- Swanson, H.L. (1993). Individual differences in working memory: A model testing and subgroup analysis of learning-disabled and skilled readers. *Intelligence*, 17, 285-332. doi: 10.1016/0160-2896(93)90003-N
- Swanson, H.L. (1996). *Swanson-cognitive processing test (S-CPT): A dynamic assessment measure*. PRO-ED, Austin, TX.
- Swanson, H. L. & Berninger, V. W. (1996). Individual differences in children's working memory and writing skill. *Journal of experimental child psychology*, 63, 358-385. doi: 10.1006/jecp.1996.0054
- Tibshirani, R., Walther, G., & Hastie, T. (2002). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63, 411-423. doi: 10.1111/1467-9868.00293
- Tiffin, J. (1968). *Purdue pegboard examiner manual*. Chicago: Science Research Associates.
- Tiffin, J. (2002). *Purdue pegboard test*. Lafayette Instrument, Lafayette, IN.

- Todman, J. B. & Dugard, P. (2001). *Single-Case and Small-N Experimental Designs: A Practical Guide To Randomization Tests*, Mahwah, NJ, USA: Lawrence Erlbaum.
- Torgesen, J. K., Wagner, R., & Rashotte, C. (2011). *TOWRE-2 Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.
- 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: Guilford Press.
- Troia, G. A., & Graham, S. (2002). The effectiveness of a highly explicit, teacher-directed strategy instruction routine changing the writing performance of students with learning disabilities. *Journal of Learning Disabilities*, 35, 290-305. doi: 10.1177/00222194020350040101
- Vanderberg, R., & Swanson, H. L. (2007). Which components of working memory are important in the writing process? *Reading and Writing*, 20, 721-752. doi: 10.1007/s11145-006-9046-6
- Van Der Steen, S., Steenbeek, H., Van Dijk, M., & Van Geert, P. (2014). A process approach to children's understanding of scientific concepts: A longitudinal case study. *Learning and Individual Differences*, 30, 84-91. doi: 10.1016/j.lindif.2013.12.004
- Van Waes, L., Leijten, M., Wengelin, A., & Lindgren, E. (2012). Logging tools to study digital writing processes. In: V. W. Berninger (Ed.) *Past, Present, and Future Contributions of Cognitive Writing Research to Cognitive Psychology*. New York: Psychology Press, Taylor & Francis Group.
- Van Waes, L., & Schellens, P. J. (2003). Writing profiles: The effect of the writing mode on pausing and revision patterns of experienced writers. *Journal of pragmatics*, 35, 829-853. doi: 10.1016/S0378-2166(02)00121-2

Wechsler, D. (2008). Wechsler Adult Intelligence Scale-Third Edition (WAIS-IV). Pearson, San Antonio, TX.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson III tests of achievement. Itasca, IL: Riverside Publishing.

Table 1

Repeated measures statistics for written output variables

Variable	Mean Pen ( <i>SD</i> )	Mean keyboard ( <i>SD</i> )	<i>F</i> -value	Sig	<i>Partial</i> $\eta^2$
Lectical analysis score (LAS)	11.16 (.21)	11.27 (.26)	8.39	.025	.14
Total writing time (min)	24.84 (9.28)	26.13 (9.97)	1.28	.263	-
Total number of words	441.83 (179.8)	577.5 (232.99)	31.59	< .001	.37
Words per minute	18.37 (4.39)	23.57 (7.72)	38.38	< .001	.42
Ratio of pause/writing time	.34 (.12)	.43 (.12)	24.29	< .001	.31

Note. Significance level adjusted using the Bonferroni correction (p-value multiplied by the number of comparisons, i.e. 5).

Table 2

Test values of writing indicators for each cluster and accompanying effect sizes

	Cluster 1		Cluster 2		Cluster 3	
	Test value	<i>d</i>	Test value	<i>d</i>	Test value	<i>d</i>
LAS pen	4.48**	1.56	-1.64	-	-3.38**	1.16
LAS keyboard	3.51**	1.1	-.04	-	-3.87**	1.39
Writing time pen (min.)	4.56**	1.6	.58	-	-5.64**	2.76
Writing time keyboard (min.)	3**	.9	1.2*	.43	-4.51**	1.76
Total words pen	5.78**	2.64	-2.26	-	-4.22**	1.59
Total words keyboard	4.87**	1.8	-2.33*	.88	-3.14**	1.06
Words/min. pen	1.52*	.42	-4.76**	2.27	2.97**	.99
Words/min. keyboard	1.26	-	-3.89**	1.63	2.4**	.77
Ratio pause/writing pen	-1.89*	.35	5.73**	3.35	-3.49**	1.22
Ratio pause/writing keyboard	-1.08	-	2.4**	.88	-1.14	-

Note. \* $p < 0.05$ ; \*\* $p < 0.01$  (p-value derived from Monte Carlo permutation tests). Positive test values indicate that the cluster scores higher than the other two clusters combined, whereas negative test values indicate that this cluster scores lower than the other two clusters combined. Means and standard deviations of the writing output measures can be found in Table 3.

Table 3

## Within-cluster modality differences and significance

	Cluster 1				Cluster 2				Cluster 3				Overall	
	Mean Pen (SD)	Mean Keyboard (SD)	<i>p</i>	<i>d</i>	Mean pen (SD)	Mean keyboard (SD)	<i>p</i>	<i>d</i>	Mean Pen (SD)	Mean keyboard (SD)	<i>p</i>	<i>d</i>	Mean pen (SD)	Mean keyboard (SD)
Lectical analysis score (LAS)	11.30 (.14)	11.40 (.21)	.02	.59	11.08 (.15)	11.26 (.25)	.02	.89	11.0 (.22)	11.04 (.21)	.31	-	11.16 (.21)	11.27 (.26)
Total writing time (min)	31.09 (5.99)	30.56 (7.7)	.63	-	26.09 (5.52)	28.91 (9.03)	.16	-	13.25 (4.56)	16.16 (7.07)	.09	-	24.84 (9.28)	26.13 (9.97)
Total number of words	595.64 (131.9)	745.24 (211.3)	< .01	.85	347.43 (67.9)	451.5 (121.6)	< .01	1.06	273.60 (82.6)	415.5 (145.0)	< .01	1.2	441.8 (179.8)	577.5 (233.0)
Words per minute	19.35 (3.24)	25.01 (6.24)	< .01	1.14	13.52 (2.10)	16.61 (5.41)	.03	.75	21.25 (4.08)	27.68 (7.82)	< .01	1.03	18.37 (4.39)	23.57 (7.72)
Ratio of pause/writing time	.31 (.07)	.42 (.12)	< .01	1.16	.49 (.05)	.50 (.12)	.38	-	.25 (.08)	.40 (.11)	< .01	1.57	.34 (.12)	.43 (.12)

Note. P-value derived from Monte Carlo permutation tests.



Table 4

## Statistics of Working Memory Measures (Mean and SD) for Each Cluster

WM measure	Overall	Cluster 1	Cluster 2	Cluster 3	$p_{\text{Cluster 1>2>3}}$	$d_{\text{Cluster 1>2>3}}$
	Mean ( <i>SD</i> )	Mean ( <i>SD</i> )	Mean ( <i>SD</i> )	Mean ( <i>SD</i> )		
Digit Backward Span (out of 14)	8.46 (2.55)	8.92 (2.86)	8.29 (2.2)	7.87 (2.33)	.02	.65
Oral Sentence Span (out of 25)	17.43 (3.59)	18.48 (3.29)	16.57 (3.52)	16.47 (3.87)	< .01	1.91

Note. P-values derived from Monte Carlo permutation tests.