JOURNAL OF LEARNING ANALYTICS

Volume 6(3), 90–104. http://dx.doi.org/10.18608/jla.2019.63.11

Detecting Contract Cheating Using Learning Analytics

Kelly Trezise[1], Tracii Ryan[2], Paula de Barba[3], Gregor Kennedy[4]

Abstract
Keystroke logging and clickstream data, both emergent areas of study in the field of learning analytics, present promising alternative methods of detecting and preventing contract cheating. The current study examines whether analysis of keystroke and clickstream data can detect when a student is creating their own authentic writing or transcribing from another source. Participants were 62 university students (47 women, 15 men) who completed three writing tasks under experimental conditions: free writing, general transcription, and self-transcription. Analyses revealed that while completing the free-writing task, participants typed in shorter bursts with longer pauses and typed more slowly with more revisions compared to the two transcription tasks. Model-based clustering was able to accurately distinguish the free-writing task from the two transcription tasks based on patterns of bursts and writing speed. Overall, these results suggest that keystroke and clickstream analysis may be able to distinguish between a student writing an authentic piece of work and one transcribing a completed work. These findings signal significant implications for the detection of contract cheating.

Notes for Practice

- Contract cheating is a significant issue for the higher education sector. There is a strong need for reliable methods to detect or prevent contract cheating. Using keystroke logging and clickstream data shows potential for this purpose.
- The current study examined whether it is possible to detect whether a student is creating their own work or transcribing from another source.
- The results indicate that it is possible to use patterns of writing activity and speed to accurately distinguish between students creating authentic writing and transcribing from another source.
- These findings point to a promising approach for detecting contract cheating and plagiarism when students transcribe work from another source.

Keywords
Keystroke logs, writing process, contract cheating, plagiarism, latent profile analysis.

Submitted: 23.10.2018 — **Accepted:** 23.07.2019 — **Published:** 13.12.2019

Corresponding author [1] Email: trezise@uchicago.edu Address: Department of Comparative Human Development, University of Chicago, 1101 E 58th St, Chicago, IL 60637, USA
[2] Address: University of Melbourne Centre for the Study of Higher Education, Elisabeth Murdoch Building, Parkville VIC 3052
[3] Address: University of Melbourne Centre for the Study of Higher Education, Elisabeth Murdoch Building, Parkville VIC 3052
[4] Address: University of Melbourne, Centre for the Study of Higher Education, Elisabeth Murdoch Building, Parkville VIC 3052

1. Introduction

Essay plagiarism, which involves copying content from other sources, sharing completed essays with other students for them to copy, and contract cheating, is a serious form of academic misconduct affecting universities worldwide (Bretag et al., 2018; Bretag et al., 2014; Davis, Grover, Becker, & McGregor, 1992; Newton, 2016). Essay plagiarism undermines the value of degrees and lowers the readiness of graduates for their chosen professions (Gallant, Binkin, & Donohue, 2015; Gullifer & Tyson, 2010). As such, postsecondary institutions have adopted systems such as Turnitin to help detect instances of plagiarism. These systems compare student-submitted essays to those of others in the class and sources on the internet to determine indices of originality. While these originality systems can be used to identify copying text from other sources or essay sharing between peers, they cannot be used to identify contract cheating (Bretag et al., 2018).

Contract cheating, which involves paying someone else to complete an assignment without attribution or recognition, has become popular with the rise of the internet (Bretag et al., 2018). There are now various dedicated businesses and websites that support this type of cheating, such as essay mills, file-sharing programs, and auction sites (Walker & Townley, 2012). Although contract cheating cannot be detected using existing originality systems, recent research suggests that keystroke logging may be a promising alternative method for ascertaining the authenticity of students' written text during the creation process (Schneider, Bernstein, vom Brocke, Damevski, & Shepherd, 2017; Usoof & Lindgren, 2008). Keystroke logging has been used to characterize the writing process (Chan, 2017; Leijten & Van Waes, 2013; Xu & Qi, 2017), but, as yet, there has been little application of this research to identify contract cheating. Recently, online writing software tools that record real-time learning analytics as keystroke logging have been implemented in higher education courses across Australia (e.g., Trezise et al., 2017. In this article, we examine whether characterization of the writing process using keystroke logging from one of these online writing software tools (Cadmus — http://cadmus.io) can be used to distinguish between instances of free writing and instances where participants were copying from other sources.

1.1. The Problem of Contract Cheating

Contract cheating is a significant issue facing the higher education sector and one that appears to be rising over time. For example, Newton (2018) conducted a systematic review of 65 studies that had measured the prevalence of contract cheating since 1978. The results indicated that an average of 3.5% of higher education students worldwide admitted resorting to contract cheating. However, when only studies conducted between 2014 and 2018 were taken into account, this prevalence rate rose to 15.7%. Worryingly, Newton (2018) argues that the actual prevalence rates are likely to be much higher, as many of the included studies had low response rates, relied on convenience samples, and failed to inform respondents that their results would be anonymous. Moreover, a study that presented students with hypothetical discrete choice experiments found that 50% of the sample would willingly resort to contract cheating given that they considered the level of cost, grade increase, risk, and penalty to be worthwhile (Rigby, Burton, Balcombe, Bateman, & Mulatu, 2015).

As noted above, contract cheating is extremely difficult to detect and eliminate (Walker & Townley, 2012). While ensuring that students have a deep understanding and appreciation of academic integrity (e.g., through educative intervention) can eliminate unintentional acts of plagiarism (Davis et al., 1992; Newton, 2016), these approaches are ineffective when students plagiarize on purpose (Devlin & Gray, 2007), as is the case with contract cheating. Furthermore, current plagiarism detection software (e.g., Turnitin) is not capable of detecting contract cheating, especially when students commission bespoke essays that are well written and referenced (Walker & Townley, 2012). Recently, Turnitin (2018) has developed a new application called Authorship Investigate, which analyzes document sets from a single author to determine similarities and differences on a number of metrics, including readability, style, and linguistic features. However, this system cannot accurately detect contract cheating; it can only flag potential instances where the essay does not appear to match other work from the same author. An alternative approach to detection could involve the use of software to automatically monitor contract cheating requests on popular websites (Clarke & Lancaster, 2006), but this approach is only useful when students make publicly accessible requests for assistance on specific assessment tasks that can be traced back to the university in question. Arguably, students and service providers would quickly adapt to such efforts by making their interactions private and inaccessible to detection. As such, there is a strong need for alternative approaches to more easily detect when students are resorting to contract cheating.

1.2. Using Learning Analytics to Detect Contract Cheating

Recent advances in online word-processing technology allow for an innovative approach to detecting contract cheating: monitoring students' writing creation processes. Software tools such as Cadmus (Trezise et al., 2017) and Inputlog (Leijten & Van Waes, 2013) record students' keystroke logs, and recent research suggests that this could be a promising alternative method for ascertaining the authenticity of students' written text during the creation process (Schneider et al., 2017; Usoof & Lindgren, 2008). For example, Schneider et al. (2017) argue that keystroke logs could offer a superior method of detecting when contract cheating has occurred, as avoiding detection of cheating by the loggers would require much more effort by the student. Students would need to transcribe the work to ensure the keystrokes had been logged and work to emulate the patterns of keystroke activity involved in natural writing processes, such as appropriate periods of keystroke activity, pauses, deletions, and revisions.

Previous research has established that keystroke logging can be useful for understanding the writing processes involved in natural essay production (Chan, 2017; Leijten & Van Waes, 2013; Xu & Qi, 2017). This technique involves recording timestamps, key presses, and pauses made by students as they create text-based work on a computer (Leijten & Van Waes, 2013). It is potentially possible to analyze keystroke logs to determine whether the processes used by the student (i.e., periods of writing generation, revision, and pauses) match those involved in authentic essay creation. Keystroke logging allows analysis of the fluency and flow of writing, the length and frequency of pauses, and patterns of revision behaviour. Using these data, it is possible to draw conclusions about students' underlying cognitive processes (Chan, 2017; Leijten & Van Waes, 2013).

One problem with relying on keystroke data alone is the absence of a link between keystroke patterns and the text itself. Due to its complex nature, the process of determining features such as the number of words written or deleted with keystroke logging can be difficult. Although keystroke logging provides excellent simple data for timing and number of keystrokes, it is more difficult to use this method to track higher processes, such as words written or deleted (compared to number of times the delete key was pressed). However, analysis of clickstream data enables a relatively high level description of writing processes, such as duration on task; words added, deleted, and pasted; and text features such as bold and highlighting. The use of clickstream data in some cases provides advantages over keystroke logging; for example, detection of words added or deleted rather than number of deletion keystrokes can reduce the effect of typos during translation, providing a simpler measure of revision behaviours. Identifying revision behaviours is key to detecting contract cheating in online word-processing software because transcription of a text (such as the purchased essay) is less likely to require lengthy revisions. Consequently, there may be benefits to using clickstream data in combination with keystroke logging to detect contract cheating.

1.3. Linking Learning Analytics Data and Writing Processes

Understanding the cognitive process involved in writing can assist in interpreting both keystroke logging and clickstream data. Essay writing is known to involve three main cognitive phases: planning, translating, and revising (Hayes, 2000; Hayes & Flower, 1986). Planning involves the production and organization of ideas, translating is the act of generating words in text to convey those ideas, and revision occurs when the author refines and edits the text they have produced. The use of these cognitive phases is thought to differ between writing tasks (Conijn, van der Loo, & van Zaanen, 2018). Transcription largely requires copying text without production or organization of ideas and little refining or editing. Consequently, compared to free writing, transcription is thought to require less planning and revision behaviour, with a greater proportion of translation.

Each of the cognitive phases involved in writing is thought to be associated with certain patterns of keystroke behaviour. Planning behaviour is associated with pauses in writing activity. Baaijen, Galbraith, and de Glopper (2012) found that writers who took long pauses were more likely to follow them with short clean bursts of new writing. They consequently proposed that long pauses may reflect planning. Xu and Qi (2017) also considered long pauses to be indicative of planning, particularly at the earlier stages of writing. Translation is associated with addition of words and keystrokes. Deane (2014) notes that translation is likely to involve longer sequences of pause-free keystrokes than revision. As such, he operationalized translation as periods of fluid typing with no pauses longer than two standard deviations from the mean pause time. Revision behaviour is associated with revising, frequency of text modifications, and deletions (Baaijen et al., 2012). Conijn et al. (2018) defined revisions as "all consecutive keystrokes where the next keystroke resulted in a lower document length, i.e., something was removed" (p. 3). Deane (2014) suggested that lengthy periods where the backspace key was used likely reflect revision. Although certain keystroke patterns are associated with different writing processes, the distinction between processes is not clear cut (Conijn et al., 2018). For example, the addition of keystrokes may reflect either translation or revision, while pauses between keystrokes may be indicative of either planning or revision (Deane, 2014). As such, researchers have aimed to determine whether it is possible to use distinct patterns of pauses and keystrokes to identify the different writing phases.

Studies seeking to understand keystroke behaviour during writing have typically sought to characterize the writing process during a free-writing task compared to a transcription task (e.g., Conijn et al., 2018). The difference between free writing and transcription is thought to represent planning and revision — the two writing methods would be expected to require equal translation. However, a fundamental confounder underpinning this assumption is that the text itself may significantly vary between participants' own writing and their transcription task. Is it possible that keystroke patterns may be affected by (1) text features, such as word length, sentence length, and "non-word" features (such as parentheses), or (2) cognitive processes related to the text, such as language fluency or text familiarity. If so, estimates of the keystroke patterns underlying all three writing processes may be affected not only by writing task but also by text complexity and familiarity. Thus, previous attempts to distinguish between free writing and transcription may be biased.

1.4. The Current Study

While there may be potential for keystroke logs and clickstream data to be used to prevent contract cheating, research is needed to define how to accurately distinguish between free-writing processes and the processes used in transcription (Baaijen et al., 2012). The current study examined whether the analysis of patterns of writing, pause, and revision activity using both keystroke logs and clickstream data can be used to distinguish between students' writing tasks (i.e., creating their own writing compared to transcribing from another source).

To examine whether typing patterns differed between writing tasks, participants' keystroke logs and clickstream data were recorded while they completed three writing tasks: free writing, general transcription of an unfamiliar text, and a self-transcription task. In the free-writing task, participants were asked to write an original essay on a specific theme. For the general transcription task, participants were given a printed essay and asked to transcribe it. The self-transcription task involved participants transcribing a printed copy of the essay they wrote in the free-writing task. This addition of the self-transcription task is a novel contribution to the literature and allows the comparison of writing patterns when (a) the text is identical but the task differs (free writing compared to self-transcription), (b) the text varies but the task is constant (general transcription compared to self-transcription), and (c) the normal examination of text and task variation (free writing compared to general transcription). For this reason, the research is largely exploratory: typing patterns for self-transcription, in comparison to both general transcription and free writing, are not yet characterized. Although this is an exploratory study, we suggest that similarities between the general transcription and self-transcription tasks might be interpreted as typing patterns attributed to transcribing, while similarities between the free-writing and self-transcription tasks might be interpreted as typing patterns attributed to features of the text.

Based on previous research focusing on keystroke logging and writing processes, it was expected that an analysis of keystroke patterns would provide insight into students' planning, translating, and revising behaviour as they completed writing tasks. In line with previous research, we expected that free writing would show greater planning and revision and less translation than the general transcription task. We also hypothesized that differences between the self-transcription task, the free-writing task, and the general transcription task would provide insight into the cognitive processes required for transcription. If only task type affects typing patterns, then we would expect to see differences between the free-writing and the general and self-transcription tasks, but no difference between the general transcription and the self-transcription tasks. If only text complexity and familiarity affect typing patterns, then we would expect to see no differences between the free-writing and self-transcription tasks, but these would both be different from the general transcription tasks. However, if task type and text complexity and familiarity affect typing patterns, then we would expect to find differences between all three typing tasks. Finally, we sought to examine whether these differences in the writing process would result in the ability to use cluster analysis to identify the typing task by examining the patterns of typing speed and bursts.

Given that *planning* is associated with longer pauses, *translation* is associated with fewer pauses and more keystrokes and word additions, and *revision* is associated with pauses and word deletions, our hypotheses were as follows:

(1) Compared to general transcription, free writing would be associated with more-frequent pauses (i.e., shorter burst duration), longer pause duration, fewer keystrokes and words added, more word deletions, more typing variability, and less typing during bursts.

(2) If task type affects

(a) planning, then self-transcription would show fewer pauses (i.e., longer burst duration) and shorter pause duration compared to free writing;

(b) translation, then self-transcription would show more keystrokes and words added compared to free writing;

(c) revision, then self-transcription would show fewer deletion and more typing during bursts compared to free writing.

(3) If text complexity/familiarity affects

(a) planning, then self-transcription would show lower pause frequency (i.e., longer burst duration) and shorter pause duration compared to general transcription;

(b) translation, then self-transcription would show more keystrokes and words added compared to general transcription;

(c) revision, then self-transcription would show fewer deletions and more typing during bursts compared to general transcription.

(4) Finally, if keystroke patterns are distinct between free writing, self-transcription, and general transcription, then a cluster analysis would identify three clusters, with each data source corresponding to a different typing task.

2. Method

2.1. Participants

Participants were 62 students (76% women, 24% men) from a large research-intensive university in Melbourne, Australia. The majority of the students (63%) were enrolled in a bachelor's degree, while 2% were completing honours, 21% were completing master's degrees, and 3% were PhD candidates (11% did not specify). With regard to discipline, 42% of the students were enrolled in STEM courses, 40% were enrolled in professional courses (e.g., accounting, commerce), and 18% were enrolled in humanities and social science courses. Almost all (92%) of the participants were right-handed. More than half of the participants (66%) were from a non-English-speaking background, but almost all (97%) primarily used English-language keyboards.

2.2. Materials

Participants were asked to complete three writing tasks using Cadmus, which is an online word-processing software tool. Cadmus has most features of other word-processing software tools, such as a main body section for writing, with style editing, inserting tables and images, bolding, and highlighting, among others. The Cadmus system was developed to provide assurances of the identity and authenticity of students' written assignments, while simultaneously educating students about academic integrity. As such, Cadmus, for example, restricts the number of characters when students copy and paste and provides them with constructive feedback (e.g., "Try paraphrasing instead. Here are some examples (…)."). Cadmus continuously records the user's actions via the keyboard and takes a snapshot of the document every two minutes, creating a temporal log of participants' writing activities.

2.3. Procedure

Ethics committee approval was obtained from the university before data collection. Participants were recruited via a public advertisement to participate in a laboratory study involving writing tasks. All participants volunteered to participate and provided informed consent. Prior to commencing the writing tasks, participants were asked to complete an initial demographic and self-regulated learning questionnaire; however, these data are outside the scope of the current paper and are not reported here.

All participants completed three writing tasks in the same order: the free-writing task was first, the general transcription task was second, and the self-transcription task was third. Although counterbalancing the order of tasks would have accounted for order effects, the order used in this study was chosen for practical reasons; inserting the general transcription task between the free-writing and self-transcription tasks allowed time for the researchers to print out all of the original essays for participants to use in the self-transcription task.

At the beginning of the experiment, participants were given the following instructions for the free-writing task: "Write an essay describing your educational experience. Include your school experiences, your transition from school to university, why and when you chose to attend the [university] and study your current degree, and your experience at the [university] so far." Participants had approximately 35 minutes for this task, and they were asked to complete and submit the task using Cadmus. After the 35 minutes were up, or they had submitted their essay, they were given a short break before the general transcription task. Before the students began the general transcription task, the experimenters handed out a printed hard copy of an essay to all participants. All participants were given the same essay and were given approximately 20 minutes to transcribe it using Cadmus. Once participants had completed the general transcription task, they were asked to submit the essay via Cadmus and take another short break. While participants were completing the general transcription task, the experimenters printed out each participant's free-writing task. For the self-transcription task, the experimenters provided each participant with a printed hard copy of their original essay from the free-writing task and asked them to transcribe it using Cadmus. They were instructed to not correct any errors in the original text but rather to type the text verbatim. Participants had approximately 20 minutes for this task, but many completed it early.

2.4. Data Processing and Analyses

Learning analytics data used in the current study include the number of words added and the number of words deleted from clickstream data and patterns of bursts and pauses from keystroke data. We were particularly interested in word addition and deletion efficiency and variability from the clickstream data. To examine this, the number of words added and deleted in each two-minute "snapshot" was calculated into a ratio of words per minute. For analysis, we calculated each participant's median and interquartile range of words added/minute and deleted/minute for each task. To examine patterns of bursts and pauses, we adopted Baaijen et al.'s (2012) definition of pauses as being interruptions to keystrokes that are longer than or equal to two seconds. Consequently, keystroke logs were separated into bursts and pauses. Participants' median time of bursts and pauses was calculated. Finally, we examined activity during bursts. To do this, an index of typing and burst behaviour was created. We decided to focus on the number of keystrokes, as this would account for both additions and deletions. Consequently, we created a ratio: keystrokes per burst. We then examined whether median keystrokes per burst differed between tasks.

In the first part of the results, we describe these measures for each task, and then we compare these measures between the tasks. For clarity purposes, we present the tasks in different orders depending on the results found in each clickstream and keystroke log measure. We used the statistical software programs R and JASP for data analysis. We applied Greenhouse-Geisser when the assumption of sphericity was violated, and Bonferroni's correction for multiple comparisons. Effect sizes for ANOVAs are reported using partial eta squared (0–1 scale). Effect sizes for post hoc tests report Cohen's d, which can be interpreted in standard deviation units: higher values indicate greater distinction between the groups.

In the second part of the results, we sought to examine whether the typing task (free writing, self-transcription, or general transcription of unfamiliar text) can be identified by examining learning analytics from both clickstream and keystroke data. To do this, a clustering analytic approach was adopted to first identify patterns (i.e., clusters) of typing speed and bursts and then examine whether these identified clusters distinguish between the three typing tasks. We used a model-based cluster model using an expectation-maximization algorithm with the MCLUST (Fraley, Raftery, & Scrucca, 2016) package in R to

ISSN 1929-7750 (online).

conduct the cluster analysis. Model-based clustering has "the ability to identify and describe different learner patterns or learning pathways" (Hickendorff, Edelsbrunner, McMullen, Schneider & Trezise, 2018, p. 3). It is a person-oriented analytical approach that focuses on characterizing individual cases on the basis of response patterns, rather than traditional variable-oriented analytical approaches that describe the relationships between variables (Bergman & Magnusson, 1997; Hickendorff et al., 2018). In this case, we sought to characterize individuals' typing tasks on the basis of patterns (i.e., clusters) of typing speed and bursts.

In this analysis, rather than treat each individual as a single case, each individual is treated as three cases (once each for the free-writing, self-transcription, and general transcription tasks). We entered typing and burst variables as indicators into the model, with a total of 180 observations in the analysis (each participant was entered three times minus missing data and several extreme outliers1). Due to difficulties with internet connection to the Cadmus website, several participants had missing data: there was one missing record for the free-writing task, one for the self-transcription task, and one for the general transcription task. Individuals with missing data were excluded from the repeated measures ANOVAs but included in the cluster analysis (i.e., only the tasks with missing data were excluded, rather than the individuals).

Finally, we examined how well the clustering assigned observations to each task. We used Bayes' rule to calculate the accuracy of the model for identifying each of the typing tasks. Bayes' rule is a common metric used to assess degree of certainty. Bayes' rule is

$$P(C \mid T) = \frac{P(T \mid C) \cdot P(C)}{P(T)}.$$

For each cluster, we calculated the Bayes value, P(C|T), which represents the probability that a randomly selected observation assigned to a given cluster, C, completed the likely corresponding task, T. For example, the cluster 1 value indicates the probability that an observation assigned to cluster 1 completed the free-writing task. P(T|C) indicates the conditional probability that an observation assigned to the cluster C was from the likely corresponding task. $P(C)$ indicates the probability that any observation was assigned to the given cluster; $P(T)$ indicates the probability that any observation was from the likely corresponding task. Consequently, the accuracy of each cluster was calculated as

$$P(C \mid T) = \frac{P(T \mid C) \cdot P(C)}{P(T \mid C) \cdot P(C) + P(NT \mid C) \cdot P(NT)},$$

where $P(NT|C)$ indicates the conditional probability that an observation assigned to cluster C completed a task that was *not* the likely corresponding task, and $P(NT)$ indicates the probability that any observation is *not* from the likely corresponding task.

3. Results

3.1. Clickstream Data: Words Added and Deleted

We first examined whether text creation speed differed between tasks. A repeated measures ANOVA examined the relationship between task (free writing, self-transcription, and general transcription) and typing efficiency (words added per minute and words deleted per minute). There was a significant relation between task and typing efficiency, $F(1.81, 103.30) = 131.51$, $p < .001$, $\eta^2_p = .70$. For subsequent analyses, we examined words added per minute separately from words deleted per minute (refer to Figure 1).

Planned comparisons were then used to examine the relationship between task and words added per minute, and between task and words deleted per minute. Words added per minute differed between tasks (see Figure 1A and Table 1). Words added per minute was higher for the self-transcription task than for the free-writing task and the general transcription task. There was no statistically significant difference between the free-writing task and the general transcription task. In summary, participants showed an order of adding words/min between tasks: self-transcription > free writing = general transcription.

[1] Three individuals had very high burst durations in the self-transcription condition (>> M + 8 × SD). These high burst durations occurred because the three cases had very few pauses. (1–3 pauses), resulting in very long burst duration (e.g., 6.4 minutes long). These three were excluded because they were so extreme and had a drastic effect on the aggregate data analysis; e.g., burst duration was M = 29 seconds, SD = 23 seconds excluding the three outliers, and M = 41 seconds, SD = 55 seconds when the three outliers were included. These cases indicate that perhaps two seconds is not an appropriate pause cut-off for self-transcription tasks and that examination of brief pauses (i.e., <2 seconds) may also provide insight into the ability to distinguish self-transcription from general transcription

Table 1. Task Comparisons for Words Added per Minute

	ANOVA and post hoc tests
(Overall)	$F(2, 114) = 109.16, p < .001, \eta^2_p = .66$
Free writing × Self-transcription	$t(57) = 12.01, p < .001$, Cohen's $d = 1.58$
Free writing × General transcription	$t(57) = 13.09, p < .001$, Cohen's $d = 1.72$
Self-transcription × General transcription	$t(57) = 1.97, p = .16$, Cohen's $d = .26$

Words deleted per minute also differed between tasks (refer to Figure 1B and Table 2). Words deleted per minute was higher for the free-writing task than for the self-transcription task and the general transcription task, and for the self-transcription task than for the general transcription task. In summary, participants showed an order of deleting words/minute speed between tasks: free writing >> self-transcription > general transcription.

Table 2. Task Comparisons for Words Deleted per Minute

	ANOVA and post hoc tests
(Overall)	$F(2, 114) = 125.19, p < .001, \eta^2_p = .69$
Free writing × Self-transcription	$t(57) = 9.64, p < .001$, Cohen's $d = 1.27$
Free writing × General transcription	$t(57) = 13.07, p < .001$, Cohen's $d = 1.72$
Self-transcription × General transcription	$t(57) = 7.62, p < .001$, Cohen's $d = 1.00$

We also examined whether individuals' variability in text creation speed varied between tasks. A repeated measures ANOVA examined the relationship between task (free writing, self-transcription, and general transcription) and variability in typing efficiency (interquartile range of words added per minute, and interquartile range of words deleted per minute). There was no relation between task and typing efficiency: $F(2, 114) = 2.93$, $p = .058$, $\eta^2_p = .05$. There was a significant difference between variability in words added and deleted: $F(1, 57) = 146.16$, $p < .001$, $\eta^2_p = .72$. We then examined variability in words added per minute separate from variability in words deleted per minute.

Planned comparisons were then used to examine the relationship between task and individual variability in words added per minute and between task and individual variability in words deleted per minute. Individual variability in words added per minute differed between tasks (see Figure 1C and Table 3). Variability in words added per minute was higher for the free-writing task than for the general transcription task, but not the self-transcription task. There was no statistically significant difference between the self-transcription task and the general transcription task. In summary, participants showed an order of variability in words added per minute: free writing > general transcription = self-transcription.

Table 3. Task Comparisons for Variability in Words Added per Minute

	ANOVA and post hoc tests
(Overall)	$F(2, 114) = 6.91, p < .001, \eta^2_p = .14$
Free writing × Self-transcription	$t(57) = 2.14, p = .110$, Cohen's $d = .28$
Free writing × General transcription	$t(57) = 4.63, p < .001$, Cohen's $d = .61$
Self-transcription × General transcription	$t(57) = 2.13, p = .11$, Cohen's $d = .28$

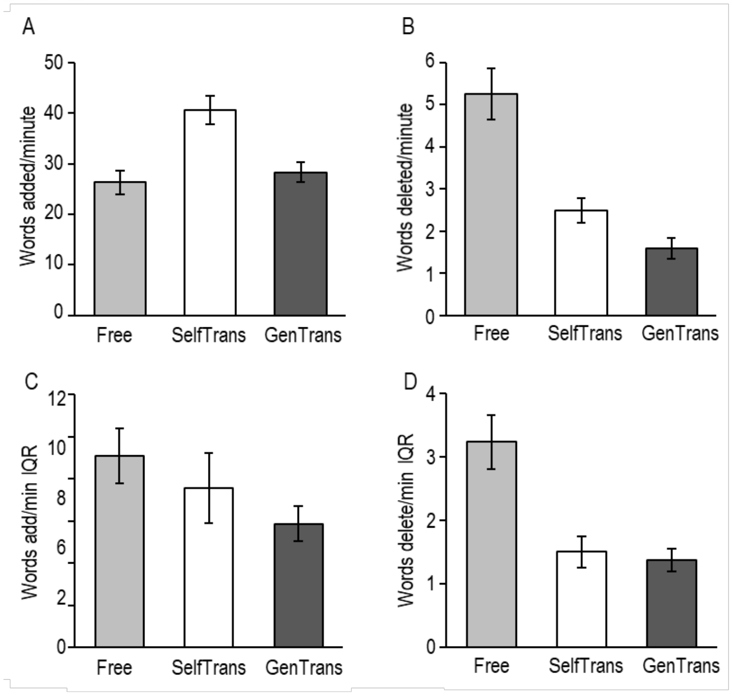

Figure 1. (A) Mean words added/minute, (B) mean words deleted/minute, (C) words added/minute interquartile range, and (D) words deleted/minute interquartile range for each task type (free writing, self-transcription, and general transcription). Error bars indicate 2 × SE.

Individual variability in words deleted per minute also differed between tasks (refer to Figure 1D and Table 4). Variability in words deleted per minute was higher for the free-writing task than for the self-transcription task and the general transcription task. There was no difference between the self-transcription task and the general transcription task. In summary, participants showed an order of variability in words deleted per minute: free writing >> self-transcription = general transcription.

Table 4. Task Comparisons for Variability in Words Deleted per Minute

	ANOVA and post hoc tests
(Overall)	$F(2, 114) = 62.61$, $p < .001$, $\eta^2_p = .52$
Free writing × Self-transcription	$t(57) = 8.72$, $p < .001$, Cohen's $d = 1.14$
Free writing × General transcription	$t(57) = 8.43$, $p < .001$, Cohen's $d = 1.11$
Self-transcription × General transcription	$t(57) = 1.09$, $p = .85$, Cohen's $d = .14$

3.2. Keystroke Data: Writing Bursts and Pauses

Next, we examined whether patterns of bursts and pauses varied between the three writing tasks. Three subjects had very high burst durations for the self-transcription task, which corresponded to very few bursts (for example, one participant had two bursts, with median burst duration of 311 seconds). These three cases were excluded from the burst duration analysis.

We examined whether burst and pause duration differed between tasks. A repeated measures ANOVA examined the relationship between task (free writing, self-transcription, and general transcription) and patterns of bursts and pauses (median number of bursts and median number of pauses). There was a significant relation between task type and burst-pause duration, $F(1.30, 70.40) = 47.22$, $p < .001$, $\eta^2_p = .47$. Because of the relation, we examined pauses and bursts separately for subsequent analyses (refer to Figure 2).

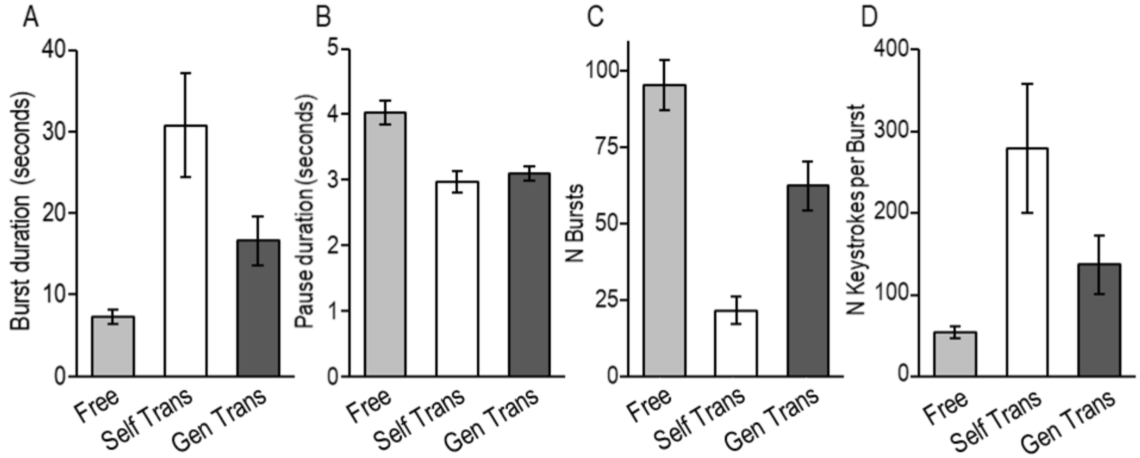

Figure 2. (A) Mean burst duration (seconds), (B) Mean pause duration, (C) Mean total number of bursts, and (D) Mean number of keystrokes per burst, for each task type (free writing, self-transcription, and general transcription). Error bars indicate 2 × SE.

Planned comparisons were then used to examine the relationship between task and burst duration, and between task and pause duration. Burst duration differed between tasks (see Figure 2A and Table 5). Burst duration was longer for the self-transcription task than for the free-writing task and the general transcription task. The general transcription task also had a longer burst duration than the free-writing task. In summary, participants showed an order of burst duration between tasks: self-transcription >> general transcription > free writing.

Table 5. Task Comparisons for Burst Duration

	ANOVA/*t*-test
(Overall)	$F(2, 108) = 44.21, p < .001, \eta^2_p = .45$
Free writing × Self-transcription	$t(54) = 7.36, p < .001$, Cohen's $d = .99$
Free writing × General transcription	$t(54) = 6.36, p < .001$, Cohen's d =.86
Self-transcription × General transcription	$t(54) = 5.49, p < .001$, Cohen's $d = .74$

Pause duration also differed between tasks (refer to Figure 2B and Table 6). Pause duration was longer for the free-writing task than for the self-transcription task and the general transcription task. There was no difference between the self-transcription task and the general transcription task. In summary, participants showed an order of pause duration between tasks: free writing > self-transcription = general transcription.

Table 6. Task Comparisons for Pause Duration

	ANOVA and post hoc tests
(Overall)	$F(2, 114) = 62.48, p < .001, \eta^2_p = .52$
Free writing × Self-transcription	$t(57) = 8.46, p < .001$, Cohen's $d = 1.11$
Free writing × General transcription	$t(57) = 9.07, p < .001$, Cohen's $d = 1.19$
Self-transcription × General transcription	$t(57) = 1.72, p = .27$, Cohen's $d = .23$

We then explored whether the number of bursts differed between tasks. A repeated measures ANOVA was used to examine the relationship between number of bursts and typing task (free writing, self-transcription, and general transcription). There was a significant relation between number of bursts and task (refer to Figure 2C and Table 7). Post hoc comparisons showed a greater number of typing bursts for the free-writing task than for the self-transcription task and the general transcription task. There were also more bursts for the general transcription task than for the self-transcription task. In summary, participants showed an order of the number of typing bursts: free writing >> general transcription > self-transcription. However, caution should be used in interpreting this finding, as participants were also given more time to complete the free-writing task.

Table 7. Task Comparisons for Number of Bursts

	ANOVA and post hoc tests
(Overall)	$F(1.43, 81.45) = 137.84$, $p < .001$, $\eta^2_p = .71$
Free writing × Self-transcription	$t(57) = 16.46$, $p < .001$, Cohen's $d = 2.16$
Free writing × General transcription	$t(57) = 5.96$, $p < .001$, Cohen's $d = .78$
Self-transcription × General transcription	$t(57) = 13.73$, $p < .001$, Cohen's $d = 1.80$

A repeated measures ANOVA examined the relationship between task (free writing, self-transcription, and general transcription), and keystrokes per burst. There was a significant relation between task and keystrokes per burst (refer to Figure 2D and Table 8). Post hoc comparisons showed a higher ratio of keystrokes per burst for the self-transcription task than for the free-writing task and the general transcription task. There was also a higher ratio of keystrokes per burst for the general transcription task than for the free-writing task. In summary, participants showed an order of the number of keystrokes per burst: free writing >> general transcription > self-transcription.

Table 8. Task Comparisons for Keystrokes per Burst

	ANOVA and post hoc tests
(Overall)	$F(1.31, 75.75) = 25.37$, $p < .001$, $\eta^2_p = .30$
Free writing × Self-transcription	$t(57) = 5.78$, $p < .001$, Cohen's $d = .75$
Free writing × General transcription	$t(57) = 4.84$, $p < .001$, Cohen's $d = .63$
Self-transcription × General transcription	$t(57) = 4.01$, $p < .001$, Cohen's $d = .52$

In the above analyses, the three typing tasks showed distinct patterns in typing speed and patterns of typing bursts and pauses. These analyses indicate that creation of free-writing text can be characterized by slower typing with more frequent errors (or evidence of revision), short typing bursts, longer pauses, and an overall greater number of typing bursts. Self-transcription can be characterized by fast typing with some errors, long typing bursts, briefer pauses, and very few typing bursts. And transcribing unfamiliar text (general transcription) can be characterized by slower typing with few errors (or little revision), moderate typing burst duration, briefer pauses, and a moderate number of typing bursts.

3.3. Cluster Analysis

Given that the above results suggest distinctive patterns in both clickstream and keystroke data, we set out to test how distinct these patterns are. In the next analytical step, we sought to examine whether the typing task (free writing, self-transcription, and general transcription) can be identified by examining patterns of words added and deleted, writing bursts and pauses, and typing during bursts. To do this, we adopted a clustering analytic approach to first identify patterns (i.e., clusters) of these metrics and then examine whether these identified clusters distinguish between the three typing tasks. Participants' three essay entries were treated as independent observations, rather than nested. In doing so, we sought to examine whether task differences occur without accounting for individual differences, such as typing skill and style.

Three measures of clickstream data and three measures of keystroke data were entered as indicator variables. The indicators were median words added per minute, median words deleted per minute, and individual interquartile range of words deleted per minute from clickstream data; and median burst duration, median pause duration, and keystroke per burst ratio from keystroke data. Individual variability of words added per minute was not included because it had a small effect size with typing task (see above analysis). Number of bursts was not included in the analysis because it was likely confounded with time on task.

The analysis is blind to the typing task. Consequently, 180 cases were entered in the analysis (each participant was entered three times — free writing, self-transcription, and general transcription — minus missing data and outliers). We chose to conduct a three-cluster analysis, to model the objective of identifying the three typing tasks.

Output from the three-cluster solution shows that the cluster sizes are .334 for cluster 1, .237 for cluster 2, and .429 for cluster 3. In contrast, perfect identification of the tasks would show a prevalence of .333 for all clusters.

Means for the three identified clusters are shown in Table 9. In addition, Figure 3 shows the density plots for the three identified clusters. Together, these show that cluster 1 is characterized by very few keystrokes per burst, short burst lengths, and medium to long pause durations. The density distributions indicate little variation between participants in the keystrokes per burst and burst length. Cluster 1 is also characterized by fewer words added per minute, more words deleted per minute, and great intra-individual variation in the number of words deleted per minute.

ISSN 1929-7750 (online).

Cluster 2 is characterized by a high proportion of keystrokes per burst, with considerable between-subjects variability. There is also a tendency for longer burst durations, but with a large amount of between-subjects variability. Pauses are of short to medium duration. Cluster 2 also shows a higher number of words added per minute, few to moderate words deleted per minute, and mostly low intra-individual variation in the proportion of words deleted per minute.

Cluster 3 is characterized by few to moderate keystrokes per burst and short to medium burst duration. Pauses are of short to medium duration. Cluster 3 also shows a moderate proportion of words added per minute. There are few words deleted per minute and very little intra-individual variation in the proportion of words deleted per minute.

Table 9. Means of the Three Identified Clusters for Each of the Indicator Variables

	Cluster 1	Cluster 2	Cluster 3
Keystrokes per burst	53.95	303.48	102.84
Burst duration	7.28	39.99	14.29
Pause duration	4.01	2.96	3.06
Words added per minute	25.93	42.04	29.25
Words deleted per minute	5.28	2.66	1.66
Words deleted per minute IQR	3.31	1.58	1.35
Most corresponding task	Free writing	Self-transcription	General transcription

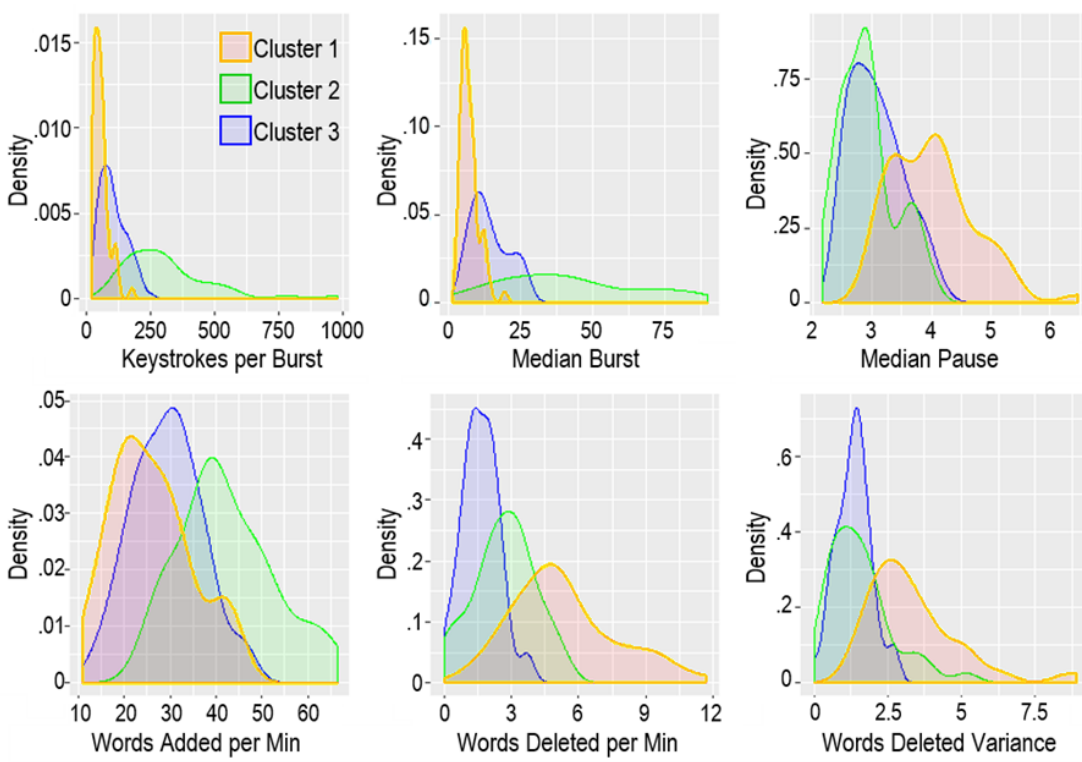

Figure 3. Density plots of the three clusters identified for each of the typing speed and burst indicators used in the cluster model.

In addition to describing the means, Figure 3 is useful for examining how each of the indicators distinguishes between the three clusters. Keystrokes per burst is useful for distinguishing between cluster 1 and cluster 2; median burst shows a similar pattern but also helps distinguish between clusters 1 and 3. Median pause length, words deleted per minute, and inter-individual variation in the words deleted per minute are useful for distinguishing cluster 1 from clusters 2 and 3. Finally, words added per minute shows distinct but overlapping distributions between all three clusters.

Finally, we examined how well the three clusters distinguished between the three typing tasks. First, we examined the relationship between the identified typing and burst clusters and the three task types. Table 10 shows the number of data cases

ISSN 1929-7750 (online).

of each identified cluster belonging to each of the typing tasks. This shows that all but two cases identified as cluster 1 were from the free-writing data. Cluster 2 was the smallest cluster, with most cases belonging to the self-transcription typing data, and some belonging to the general transcription typing data. Cluster 3 was the largest cluster, containing almost all general transcription task data, over one-third of the self-transcription task data, and a few of the free-writing task data.

Assuming cluster 1 represented the free-writing task, cluster 2 represented the self-transcription task, and cluster 3 represented the general transcription task, we then used Bayes' rule to calculate the accuracy of the model for identifying each of the typing tasks.

Cluster 1 identified the free-writing task with an accuracy of 95.04%. Cluster 2 identified the self-transcription task with an accuracy of 50.01%, and Cluster 3 identified the general transcription task with an accuracy of 85.85%. These findings show that the cluster analysis on typing speed and burst patterns was able to distinguish free writing from transcription but unable to accurately distinguish general transcription from self-transcription.

Table 10. Number of Data Cases of Each Identified Cluster Belonging to Each Typing Task

	Identified cluster		
	Cluster 1	Cluster 2	Cluster 3
Free writing	58	0	3
Self-transcription	1	31	26
General transcription	1	9	51
Probability of cluster accuracy	.95	.50	.86

4. Discussion

In this study, we sought to examine whether analysis of keystroke and clickstream data can distinguish between three distinct writing tasks: creating free-writing text; transcribing an unfamiliar piece of text; and transcribing their own, previously written text. As hypothesized, creation of free-writing text was associated with slower typing (fewer words per minute), shorter typing bursts, longer pauses, faster word deletion rate and more variability in word deletion rate, and fewer keystrokes per writing burst. We also examined whether keystroke patterns for self-transcription differ between free writing and general transcription. Our findings show keystroke patterns for the self-transcription task were highly distinct from free writing, and similar to but distinct from general transcription. Overall, findings show keystroke patterns for the free-writing task were more similar to the general transcription task than the self-transcription task. Below we further interpret the findings and discuss the possibility of keystroke logging as a method of detecting contract cheating.

We hypothesized that the keystroke patterns would differ between the free-writing task and the two transcription tasks because creating free-written text requires students to engage in planning, translating, and revising behaviours. Conversely, transcription is thought to require translating, with little revising or planning behaviours. Our findings largely support this hypothesis. Participants showed a greater number of behaviours indicating planning (e.g., longer and more frequent pauses) and revising (more words deleted per minute and variability in deletions). With a greater proportion of time taken up with planning and revising, participants have less time available for translating (fewer words added per minute). These three behaviours were represented by a smaller keystroke per burst ratio. The keystroke patterns support previous research that has characterized keystroke patterns during the writing process (Chan, 2017; Schneider et al., 2017; Usoof & Lindgren, 2008). In this study, we extend these previous works to show that keystroke patterns for typing a free-written piece of text differed to the extent that a cluster analysis was able to identify the free-written text to a 95.04% accuracy (accounting for false positives and false negatives). This finding indicates that analysis of keystroke logging can be used to distinguish between the production of free-written text and transcription of other work. Moreover, the largest differences in the study were between free writing and self-transcription, suggesting that if text is kept constant (i.e., does not differ between tasks), then keystroke differences between free writing and transcription are larger than previously thought.

Comparison of self-transcription and general transcription shows differences in some typing features. Self-transcription had significantly more words added and deleted per minute, longer but fewer bursts, and higher typing activity (i.e., number of keystrokes per burst) than general transcription. However, the cluster analysis was not successfully able to accurately distinguish between these two types of transcription. There was a degree of accuracy in classifying the self-transcription task and the general transcription task, but there was misclassification in cases of students typing faster and showing longer typing bursts.

Overall, the results of this exploratory study indicate that it may be possible to use keystroke and clickstream data to detect contract cheating. This is a promising initial outcome; however, additional research is needed to replicate and extend the findings reported here. For example, some scholars have suggested that an individual's burst and pause durations may differ according to their typing proficiency (Deane, 2014). Therefore, in future studies, it would be useful to include a measure of typing fluency (i.e., average number of words typed per minute using a standardized text). This would assist in the identification of typing speed fluctuations in each of the conditions and may help to more accurately distinguish between the phases of translation, planning, and revising. In addition, it would be interesting to use the analytical approach described in this study in conjunction with alternative methods of contract-cheating detection. One possibility involves using metrics from the Authorship Investigate application (Turnitin, 2018), such as style and linguistic features. The combination of multiple approaches would provide a valuable form of data triangulation and may contribute to the development of a highly predictive approach to detecting contract cheating.

Although not the focus of this study, examination of a general transcription task and a self-transcription task allows comparison of transcription for a constant, unfamiliar text, and a varied, familiar text. Our findings indicate that while there are task-dependent differences between self-transcription and general transcription, there also appears to be a very large participant-level effect on typing patterns. This is indicated by the cluster analysis, which saw some measures show significant within-cluster variation for the two transcription tasks. Within-group variance in keystroke patterns did not differ between the two transcription tasks. This provides insight into the interplay between text factors and individual differences in keystroke patterns. Smaller within-group variance for the general transcription would require that the task (transcription) and the text itself (e.g., paragraph/sentence/word structure) be the key "factors" affecting typing patterns. However, several factors may contribute to individual differences in the general transcription task, for example, cognitive capacity (e.g., remembering what needs to be transcribed), typing speed, duration of long bursts, and possibly affective and motivational factors. Moreover, if general transcription showed small within-group variance, it would suggest that the participants tend to type according to structures within the text itself, for example, typing bursts defined by paragraphs and brief pauses at the end of sentences or paragraphs. Future research could seek to understand the mechanisms behind these differences and further explore the usefulness of applying these keystroke logs to predictive models (e.g., decision tree, random forest, support vector machine; Heuer & Breiter, 2018) to examine whether they can distinguish between self-transcription and general transcription. This would be useful for identifying situations when students transcribe their own text that was created using another platform (e.g., Microsoft Word, Notepad), rather than online word-processing software with keystroke-logging capabilities (e.g., Cadmus).

We acknowledge that the use of writing tools that collect keystroke and clickstream data to detect contract cheating does come with constraints in relation to its implementation. In the case of this study, students were limited to using prescribed writing software (i.e., Cadmus) to produce their essays. Demanding that students use a particular software program could result, for example, in extra work if they have to copy their essay across from other software programs, or opposition when asking them to switch from their usual word processing software programs (Richman, 2016). One possible solution is to provide additional features that facilitate and better support students' writing process, making the use of a particular software program more appealing to them (Buckingham Shum et al., 2016). For example, Cadmus offers dedicated spaces for note taking and reference management, automated feedback on plagiarism, and personalized instructions for the essay, and facilitates the process for teachers to provide formative feedback to students. It has been successfully implemented in writing assessments in medium and large classes in Australian universities (e.g., Trezise et al., 2017). Another solution could involve creating add-ons so that currently available writing software programs, such as Microsoft Word and Google Docs, could collect keystroke and clickstream data. Either way, scalability would require thorough planning and scaffolding from educators on how to embed similar tools within specific learning designs (Bakharia et al., 2016).

5. Conclusion

Acts of plagiarism, particularly contract cheating, continue to be a significant issue for the higher education sector. Researchers and higher education commentators have argued that a holistic approach is required to counter contract cheating specifically, including educative approaches, alterations to learning design and assessment practices, and technical solutions (Bretag et al., 2018; Newton, 2018). The research presented in this paper indicates that, by using analytic techniques, it is possible to determine whether students are producing new writing or transcribing text. These findings provide a promising approach to detection of contract cheating that is worthy of further exploration.

Declaration of Conflicting Interest

The authors have no personal conflict of interest in undertaking this research or the preparation of the paper. The University of Melbourne has a financial interest in Vericus, the developer and vendor of Cadmus, the software tool used in this investigation.

Funding

An aspect of this research was funded by Vericus.

Acknowledgments

Thank you to the reviewers of this article for their comments and expertise.

References

Baaijen, V. M., Galbraith, D., & de Glopper, K. (2012). Keystroke analysis: Reflections on procedures and measures. *Written Communication, 29*(3), 246–277. https://dx.doi.org/10.1177/0741088312451108

Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gašević, D., Mulder, R., Williams, D., Dawson, S., & Lockyer, L. (2016). A conceptual framework linking learning design with learning analytics. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 329–338). New York: ACM. https://dx.doi.org/10.1145/2883851.2883944

Bergman, L. R., & Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology, 9*(2), 291–319. https://dx.doi.org/10.1017/S095457949700206X

Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., Rozenberg, P., …, van Haeringen, K. (2018). Contract cheating: A survey of Australian university students. *Studies in Higher Education*. https://dx.doi.org/10.1080/03075079.2018.1462788

Bretag, T., Mahmud, S., Wallace, M., Walker, R., McGowan, U., East, J., …, James, C. (2014). "Teach us how to do it properly!" An Australian academic integrity student survey. *Studies in Higher Education, 39*(7), 1150–1169. https://dx.doi.org/10.1080/03075079.2013.777406

Buckingham Shum, S., Knight, S., McNamara, D., Allen, L., Bektik, D., & Crossley, S. (2016). Critical perspectives on writing analytics. *Proceedings of the 6th International Conference on Learning Analytics & Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 481–483). New York: ACM. https://dx.doi.org/10.1145/2883851.2883854

Chan, S. (2017). Using keystroke logging to understand writers' processes on a reading-into-writing test. *Language Testing in Asia, 7*(10), 1–27. https://dx.doi.org/10.1186/s40468-017-0040-5

Clarke, R., & Lancaster, T. (2006). *Eliminating the successor to plagiarism? Identifying the usage of contract cheating sites.* Paper presented at the 2nd International Plagiarism Conference, 19–21 June 2006, Gateshead, UK. https://www.researchgate.net/publication/228367576_Eliminating_the_successor_to_plagiarism_Identifying_the_usage_of_contract_cheating_sites

Conijn, R., van der Loo, J., & van Zaanen, M. (2018). What's (not) in a keystroke? Automatic discovery of students' writing processes using keystroke logging. In A. Pardo, K. Bartimote, G. Lynch, S. Buckingham Shum, R. Ferguson, A. Merceron, & X. Ochoa (Eds.), *Companion Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (LAK '18), 5–9 March 2018, Sydney, Australia. Society for Learning Analytics Research. http://solaresearch.org/uploads/LAK18_Companion_Proceedings.pdf

Davis, S. F., Grover, C. A., Becker, A. H., & McGregor, L. N. (1992). Academic dishonesty: Prevalence, determinants, techniques, and punishments. *Teaching of Psychology, 19,* 16–20. https://dx.doi.org/10.1207/s15328023top1901_3

Deane, P. (2014). Using writing process and product features to assess writing quality and explore how those features relate to other literacy tasks. *ETS Research Report Series, 2014*(1), 1–23. https://dx.doi.org/10.1002/ets2.12002

Devlin, M., & Gray, K. (2007). In their own words: A qualitative study of the reasons Australian university students plagiarize. *Higher Education Research & Development, 26*(2), 181–198. https://doi.org/10.1080/07294360701310805

Fraley, C., Raftery, A. E., & Scrucca L. (2016). MCLUST: Gaussian Mixture Modelling for Model-Based Clustering, Classification, and Density Estimation, 2016. https://CRAN.R-project.org/package=mclust. R package version 5.2.

Gallant, T. B., Binkin, N., & Donohue, M. (2015). Students at risk for being reported for cheating. *Journal of Academic Ethics, 13*, 217–228. https://dx.doi.org/10.1007/s10805-015-9235-5

Gullifer, J., & Tyson, G. A. (2010). Exploring university students' perceptions of plagiarism: A focus group study. *Studies in Higher Education, 35*(4), 463–481. https://dx.doi.org/10.1080/03075070903096508

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*. Newark, Delaware, USA: International Reading Association.

Hayes, J. R., & Flower, L. S. (1986). Writing research and the writer. *American Psychologist, 41*(10), 1106–1113. http://dx.doi.org/10.1037/0003-066X.41.10.1106

Heuer, H., & Breiter, A. (2018). Student success prediction and the trade-off between big data and data minimization. In Krömker, D., & Schroeder, U. (Eds.), *DeLFI 2018 — Die 16. E-Learning Fachtagung Informatik*. Bonn: Gesellschaft für Informatik e.V. (pp. 219–230)

Hickendorff, M., Edelsbrunner, P., McMullen, J., Schneider, M., & Trezise, K. (2018). Informative tools for characterising individual differences in learning: Latent class, latent profile, and latent transition analyses. *Learning and Individual Differences, 66,* 4–15. https://dx.doi.org/10.1016/j.lindif.2017.11.001

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://dx.doi.org/10.1177/0741088313491692

Newton, P. (2016). Academic integrity: A quantitative study of confidence and understanding in students at the start of their higher education. *Assessment and Evaluation in Higher Education, 41*(3), 482–497. https://dx.doi.org/10.1080/02602938.2015.1024199

Newton, P. (2018). How common is commercial contract cheating in higher education and is it increasing? A systematic review. *Frontiers in Education, 3*. https://doi.org/10.3389/feduc.2018.00067

Richman, M. (2016, July 28). Microsoft Word, Google Docs and Millennials [Blog post]. https://techpinions.com/microsoft-word-google-docs-and-millennials/46660

Rigby, D., Burton, M., Balcombe, K., Bateman, I., & Mulatu, A. (2015). Contract cheating and the market in essays. *Journal of Economic Behavior & Organization, 111,* 23–37. http://dx.doi.org/10.1016/j.jebo.2014.12.019

Schneider, J., Bernstein, A., vom Brocke, J., Damevski, K., & Shepherd, D. C. (2017). Detecting plagiarism based on the creation process. *IEEE Transactions on Learning Technologies, 11*(3), 348–361. https://dx.doi.org/10.1109/TLT.2017.2720171

Trezise, K., de Barba, P. G., Jennens, D., Zarebski, A., Russo, R., & Kennedy, G. (2017). A learning analytics view of students' use of self-regulation strategies for essay writing. In H. Partridge, K. Davis, & J. Thomas. (Eds.), *Me, Us, IT! Proceedings ASCILITE2017: 34th International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education*, 4–6 December 2017, Toowoomba, Queensland, Australia (pp. 411–421).

Turnitin (2018). *Turnitin — Authorship Investigate.* Retrieved from https://www.turnitin.com/products/authorship-investigate

Usoof, H., & Lindgren, E. (2008). Who is who and doing what in distance education?: Authentication and keystroke dynamics. *Tidskrift för lärarutbildning och forskning, 3–4*, 175–187. http://umu.diva-portal.org/smash/record.jsf?pid=diva2%3A208940&dswid=4818

Walker, M., & Townley, C. (2012). Contract cheating: A new challenge for academic honesty? *Journal of Academic Ethics, 10*(1), 27–44. https://dx.doi.org/10.1007/s10805-012-9150-y

Xu, C., & Qi, Y. (2017). Analyzing pauses in computer-assisted EFL writing — A computer keystroke log perspective. *Educational Technology and Society, 20*(4), 24–34. http://www.jstor.org/stable/26229202