

A learning analytics view of students' use of self-regulation strategies for essay writing

Kelly Trezise

University of Melbourne

Paula G. de Barba

University of Melbourne

David Jennens

Vericus Pty Ltd

Alexander Zarebski

Vericus Pty Ltd

Robert Russo

Vericus Pty Ltd

Gregor Kennedy

University of Melbourne

Essay writing is a fundamental part of higher education. Students' use of self-regulatory skills, such as time management and planning and writing strategies, while writing essays predicts better writing quality. Current characterisations of the relationship between self-regulation and essay writing are limited by the difficulty of assessing self-regulation in real-life essay writing contexts. This paper reports on a novel approach to examine students' use of self-regulation strategies in a real-life setting, using learning analytics. Four case studies are presented to illustrate similarities and differences in students' use of time management, planning and writing strategies. Participants managed their time in very different ways to complete the assignment. They were active over a different number of days, engaged in sessions of different durations, and at different times of the day. The participants used variety of approaches to their writing: one participant started early and allowed editing time, another typed gradually over a number of days, and two participants waited until the due date to complete the essay, with varying amounts of editing. Findings from this research contribute to a novel detailed empirical evidence of different essay preparation behaviour in real-life settings. After further studies with a variety of essay types and student samples, there may be significant value in using the approached outlined in this paper as the basis of tools they provide students with advice and support in their essay preparation.

Introduction and background

Essay writing is a widely used type of assessment in higher education. Essay writing requires students to use high levels of self-regulation; students' use of self-regulatory skills while essay writing predicts better writing quality (Santangelo, Harris, & Graham, 2015; Zimmerman & Risemberg, 1997). However, current characterisations of the relationship between self-regulation and essay writing are limited by the difficulty of assessing self-regulation in real-life essay writing contexts. Research on essay writing processes has heavily relied on self-report methods, either after essay writing in a real-life assessment context (e.g., questionnaires, Torrance, Thomas, & Robinson, 2000), or during essay writing in a lab study (e.g., think aloud protocols, Stratman & Hamp-Lyons, 1994). Learning analytics provides a way to more precisely characterise essay writing and the use of self-regulatory processes (e.g., Azevedo, 2014; van den Bergh & Rijlaarsdam, 2013). The use of learning analytics allows examining the writing process in a less intrusive manner, and, most importantly, moves investigations from the

laboratory to real-life settings. In this paper we outline a study that used learning analytics tools to examine students' self-regulatory skills while essay writing in a real-life setting.

Models of self-regulation in writing argue students' personal processes are a key factor to their writing process (e.g., Flower & Hayes, 1981; Zimmerman & Risemberg, 1997). Zimmerman and Risemberg (1997) propose personal processes involve time planning and management, setting goals, setting self-evaluative standards to assist in monitoring performance, using cognitive strategies, and using mental imagery. This study focuses on two of the personal processes: *time planning and management*, and use of *cognitive strategies*. Time planning and management requires students to estimate and manage their time for essay writing. For example, estimating the total amount of time they will need to dedicate to essay writing, and how they will break that time into smaller blocks of time (Zimmerman & Risemberg, 1997). Cognitive strategies used in writing are



This work is made available under a [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) licence.

related to planning the structure of the essay, producing the piece of writing itself, and revising it through iterative editing (Zimmerman & Risemberg, 1997). Both processes are considered important for improving writing (e.g., Hayes & Nash, 2013; Santangelo, Harris, & Graham, 2015). However, we are yet to develop a precise characterisation of the essay writing process itself and its relation to the use of self-regulatory skills.

Although well-described, personal processes involved in essay writing (e.g., time planning and management and cognitive strategies) are rarely examined in the context of real-life essay writing. Students' writing processes are largely unobservable. Educators have access to the final version of an essay, written using paper-and-pencil or commonly used word-processing software. While teaching staff can review essays by asking students to hand in outlines or preliminary drafts, this can be unfeasible when teaching courses with large student numbers. Together, the inability to measure writing processes, combined with the time burden of reviewing drafts effectively limits educators' ability to provide students with feedback during the writing process. Recent advances in word-processing technology may present a solution to this problem for both educators and researchers. An advantage of these technological developments is that large cohorts of students' essay planning and writing strategies (e.g., self-regulation) can be analysed through learning analytics.

Learning analytics is the "measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs" (Siemens & Gašević, 2012, p.1). Learning analytics can provide educators with the opportunity to better understand and act upon students' writing process (Gašević, Dawson, & Siemens, 2015). Over the last decade, the use of learning analytics to measure students' self-regulated learning skills has taken a central role in the field (Winne, 2010). Recent research has used learning analytics to examine time students took to write characters and words during an in-class essay assessment task. For example, Deane (2014) examined writing bursts in number of characters, duration of pauses between words, and between sentences, and time spend copying and pasting text. And Eklundh and Kollberg (2013) captured students' audit logs when revising their essays using a specific software. One limitation of these studies is that they have only examined students' essay writing over a brief writing period (e.g., one hour of class, Deane, 2014). In real-life contexts, students may be given several weeks to write the essay, and will need to use time planning and management skills to write the essay. Given students' use of writing processes varies with context (Biggs, 1988; Kellogg, 1988), it is likely that writing processes required for a one-hour essay-writing

assignment will differ to those required for a multi-week assignment.

In the current study, students' use of self-regulation strategies for essay writing is examined in real-time and in a real-life context, using descriptive learning analytics. The focus of the investigation was on students' self-regulation strategies; specifically, their skills in managing time dedicated to their essay, planning essay structure, and writing their essay over a three-week period. As such, two questions underpinned the investigation. How do students manage their time to work on the essay across the assignment period? And how do students use different planning and writing strategies to complete their assignments? We use a combination of cohort data and case studies to describe and compare tertiary students' use of time management and cognitive strategies while writing an essay using an online writing platform.

Method

Participants and context

Participants were 107 students from a Business undergraduate course at the University of Melbourne. Ethics committee approval was obtained from the University and all participants provided informed consent. Participants were asked to complete a 1,000-word essay as part of their course, worth 10% of their final mark. Teaching staff marked participants' performance in this assignment using a score from 0 to 100. Participants could choose between two topics: (1) "The business of business is to make profits", and (2) "What is the business case for Corporate Social Responsibility?". Participants were instructed to include personal opinions about the material covered in the course. The aim of the essay was to assess participants' understanding of course material. Essay guidelines were released to participants on 8th of March and the due date for the essay was 27th of March. In total, participants had 19 days to work on the assignment.

Participants were instructed to complete their essay using Cadmus, an online word-processing software tool. Cadmus has most features of other word-processing software tools, such as main body section for writing with style editing, inserting tables and images, bolding, highlighting, among others. Some of the additional features of Cadmus include a section for dedicated note taking (referred to as "notes"), and a restriction of pasting at a limit of 90 words from external sources in a single paste. Cadmus is designed as a tool to support the development of participants understanding of issues associated with academic integrity. As such, a design feature is the display of automated warnings to participants related to academic integrity. For example, a warning appears if participants are close to reaching the paste limit. Participants were encouraged to use the Cadmus software for the whole essay construction:

through planning the structure of the essay, to producing the text, revising, and adding references.

Measures

Cadmus records the actions of participants via learning analytics while they work on their essay. 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. Learning analytics used in the current study include: number of sessions by student during the assignment period, duration of these sessions, number of words added and deleted, number of words copied and pasted, and which sections of the document they used (main body and notes). These measures were then organised and processed to represent participants' use of time management and cognitive strategies, such as planning and revising the essay.

Time management was represented by three variables. *Number of days* participants spent working on the essay, a count of days active in the Cadmus system. *Number of sessions* participants spent working on the essay. A session consists of a period of time greater than 30 minutes working in the Cadmus system, without interruption (Mao, Kamar, & Horvitz, 2013). *Active duration* was the amount of time a student presented any activity within a session (e.g., typing or deleting words).

Cognitive strategies for writing were divided in two categories: planning and working on the essay. Planning relates to how participants used the notes section, and was represented by four variables: *use of notes* section during the assignment, *number of added words to notes*, *deleted words from notes*, and *pasted words to notes*. The *use of notes* is a binary measure, and remaining measures are count data. Working on the essay relates to how participants used the main body section, represented by three variables: *number of typed words to main body*, *deleted words from main body*, and *pasted words to the main body*. For the class assessment, participants were marked on the contents of their main body section at time of submission.

Research design, data processing and data analysis

We used a multiple case studies design to examine similarities and differences between the cases (Baxter & Jack, 2008). Descriptive statistics were used to characterise the whole cohort, while descriptive temporal analysis was used for four case studies chosen to illustrate the different time management and writing strategies participants used.

Raw data was extracted from the Cadmus platform in CSV format for all 111 participants in the class. After initial inspection, four participants were found to have not

completed their assignment, presenting a very low final word count to their essay (below 600 words); these participants were removed from further analyses. For the descriptive temporal analysis, closer inspection revealed that 5% of the learning analytics were recorded in intervals longer than two minutes (ranging from just over 2 minutes to 17 days). This was due to technical challenges, such as interrupted internet connection. These records were substituted for the median: two minutes. Statistical softwares R and IBM SPSS were used for data analysis.

Results

The results section is presented in two parts. First, we use descriptive statistics to examine how all participants managed their time and used planning and revising cognitive strategies when working on the essay. Second, four case studies are presented to illustrate the different writing strategies participants used to work on their essay.

Descriptive statistics

Length of time that participants spent completing their essay in the Cadmus system ranged from four hours to 14 days. Participants completed their essay in an average of four and a half days ($SD=2.21$). Figure 1 displays a histogram of the number of days and sessions participants spent writing their essay. One participant completed their essay in one single session, while five participants took 20 or more sessions to complete their essay. Of the 107 participants, 23 participants did not use the notes section. Participants added an average of 2425 words to the main body of their essay ($SD=1334$), and deleted an average of 1370 words ($SD=1318$).

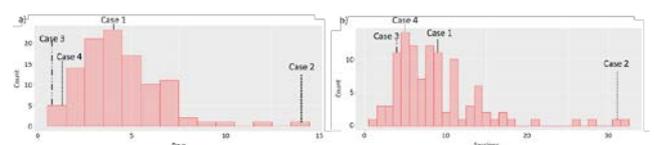


Figure 1: (A) number of days, and (B) number of sessions participants spent on the essay. Each case study is also identified

Case studies

The descriptive statistics analysis showed that participants used different strategies to complete their essay related to total of days, session duration and distribution, and writing strategies used. In this section, four case studies are presented to illustrate distinct applications of these strategies, using a descriptive temporal analysis (Figures 2 to 5). The four cases were chosen as they showed significant variation in the time management and writing strategies of interest in this investigation. The representativeness of each of these case studies is shown in Figure 1. Case 1 reflects the mode number of days. Case 2 represents the largest number of

days and number of sessions. Case 3 and 4 both reflect the smallest number of days, with Case 4 also showing the mode number of sessions. Together, these case studies represent typical and atypical examples of time management strategies.

Case study 1 – AB (Figure 2)

AB completed the essay over four days, in 9 sessions. Sessions varied from 12 minutes to 3 hours and 9 minutes, and active duration varied between 12 minutes and 1 hour and 50 minutes, with a median of 32 minutes. AB worked on the essay from 12am to 12pm across the assignment period, on the days leading up to submission day. AB did not use the notes section, and started to work on the essay directly on the main body section. The first day working on the assignment (day 16) was AB's most productive day in terms of added words. Over four sessions, AB added the highest number of words in a day, including typed and pasted words, and deleted the highest number of words in a day. On the second day working on the assignment (day 17), AB started with a large deletion, followed in the next session with a paste of similar number of words. This suggests that AB cut and pasted a large amount of text, indicating the student was restructuring their writing. On the third day working on the assignment (day 18), AB spent little time in Cadmus, with few words typed, pasted, and deleted. Moreover, the student reached the word limit for the essay. On submission day, there was a similar number of words typed and deleted, indicating revision and editing. AB's final grade was 85%.

Case study 2 – CD (Figure 3)

CD completed the essay over 14 days, in 31 sessions. Sessions varied from 2 minutes to 2 hours and 47 minutes. Active duration ranged between 2 minutes to 1 hour and 22 minutes, with a median of 6.2 minutes. CD consistently worked on the essay from 12am to 12pm across the assignment period, and mainly on weekdays (days 10 and 11 were weekend). CD made use of the notes section, but did not use the paste function while writing the essay. On the first three days working on the assignment, CD focused on the notes section. On the third day working on the essay, CD worked on the main body section, while still making a small contribution to the

notes section. For the following three days, CD made small contributions to the main body. Days 10 and 11 were the weekend, with CD working on this assignment Saturday but not on Sunday. At day 13, CD returns to adding words to the notes section. After this day, CD does not make any further contributions to the notes section. On days 14 and 15, CD showed a high level of activity in the main body, with high number of typed and deleted words, indicating some revision behaviour. After two days of rest (days 16 and 18), and a day with some contribution in a short session (day 17), the submission day arrives. CD engages in the longest session of this assignment with active duration of 1 hour and 20 minutes, and has the highest number of words typed and deleted in a day indicating both writing and revision behaviour. CD started work on the essay early, and worked on the essay over many days. For many days, CD spent little time in Cadmus (between 2 and 20 minutes). CD made use of notes, and had a high number of deleted words, indicating the student used notes to plan the essay before starting to write, and spent time revising the essay during the writing process. CD's final grade was 81%.

Case study 3 – EF (Figure 4)

EF completed the essay in one day (day 19), over four sessions. EF worked on the essay from 4am to 2pm on submission day. Sessions ranged from 2 minutes to 3 hours and 53 minutes. Active duration ranged between 2 minutes and 1 hour and 42 minutes, with a median of 8 minutes. EF started working on the notes section at 4am, typing and deleting a moderate number of words, suggesting some revision process during that writing process. Up until 10am EF made sparse contributions to the notes section. These short frequent sessions could indicate some sort of task switching between writing and reading materials, for example. From 10 am until 12pm, onwards, EF typed in the whole assignment in one long session, with active duration of 1h 42min. This session was mostly characterised by the constant addition of typed words, with some deletion of words throughout the session. At this point, EF was mainly writing down ideas, and probably revising small sections. EF did not paste any words when working on the assignment. EF's final grade was 70%.

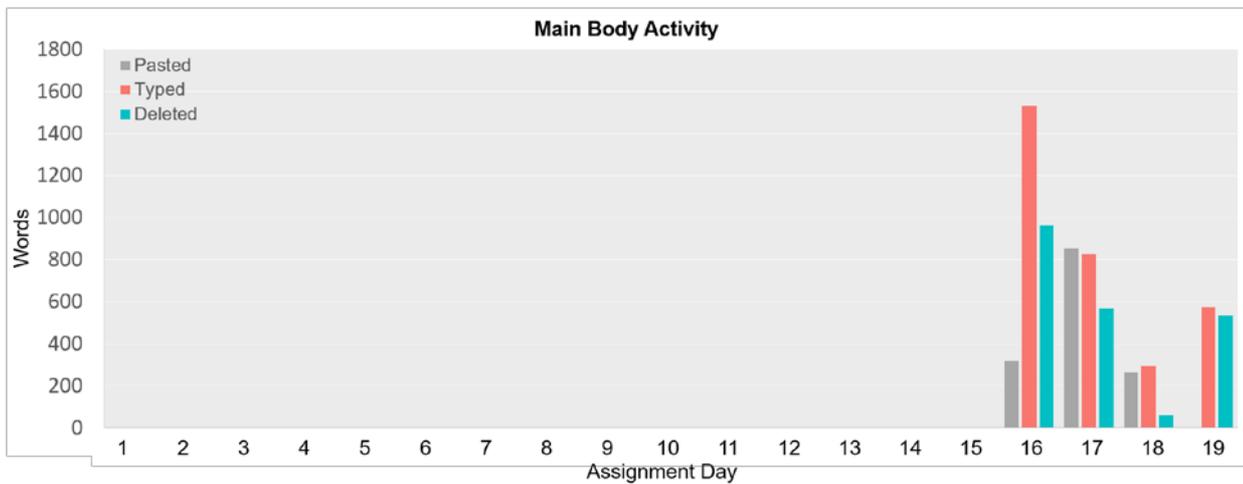
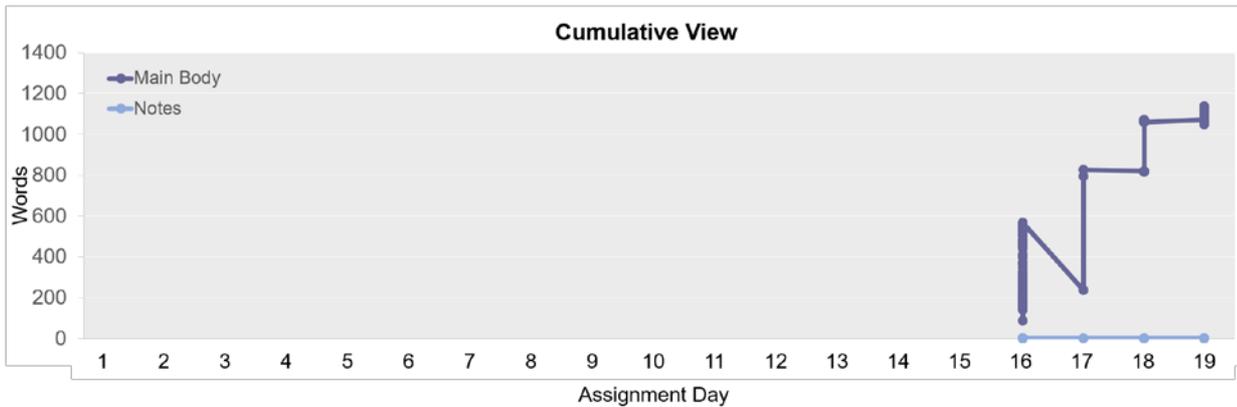
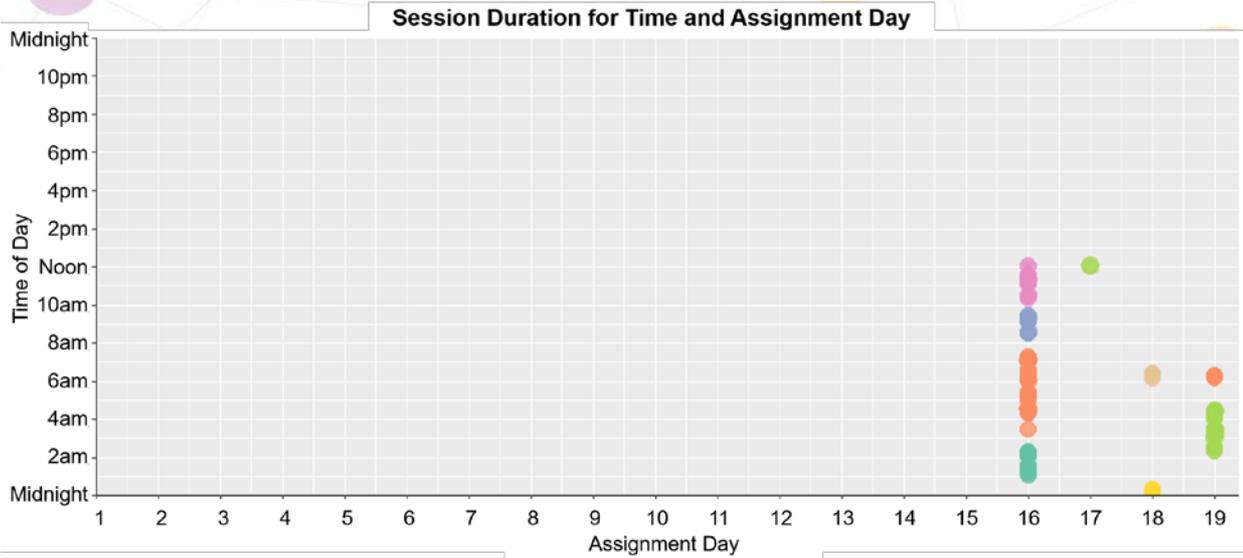


Figure 2: Case study 1: AB. Figure shows Session Duration (top) as a function of assignment day, and time of day throughout the assignment period (different colours indicates different sessions); Cumulative View (2nd from top) as a function of assignment day and number of words in the notes and main body; Main Body Activity (bottom) as a function of words typed and deleted each day.

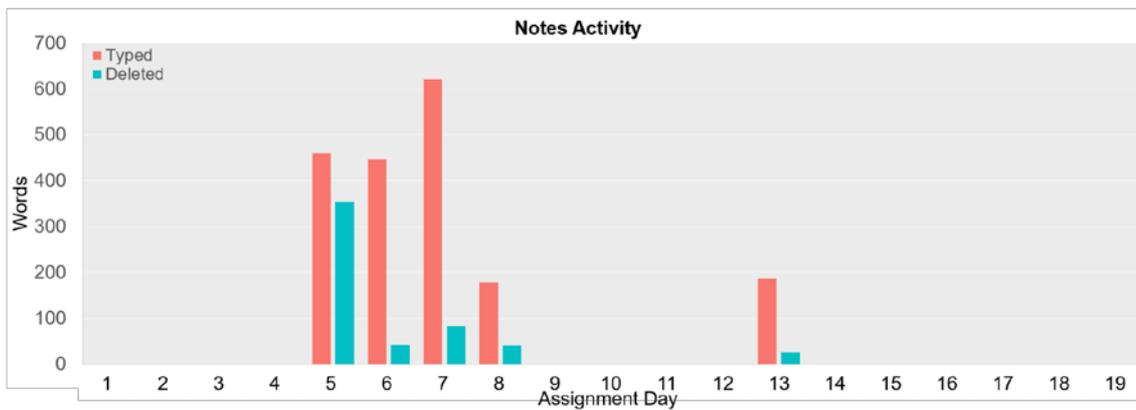
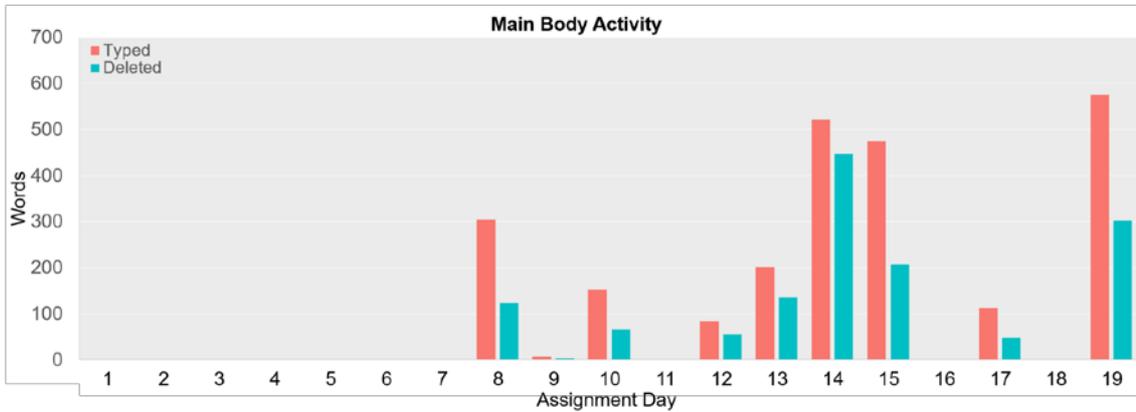
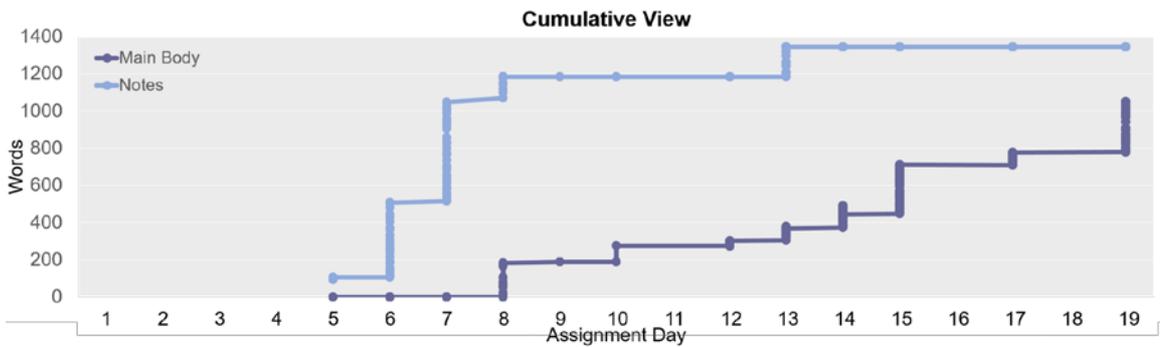
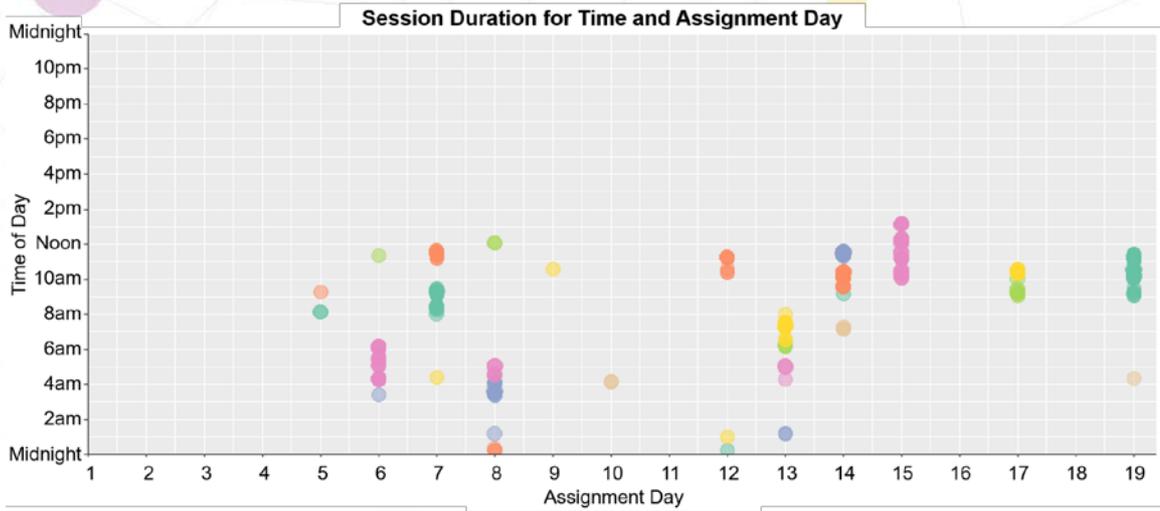


Figure 3: Case study 2: CD. Figure shows Session Duration (top) as a function of assignment day, and time of day throughout the assignment period (different colours indicates different sessions); Cumulative View (2nd from top) as a function of assignment day and number of words in the notes and main body; Main Body Activity (3rd from top) as a function of words typed and deleted each day; Notes Activity (bottom) as a function of words typed and deleted each day.

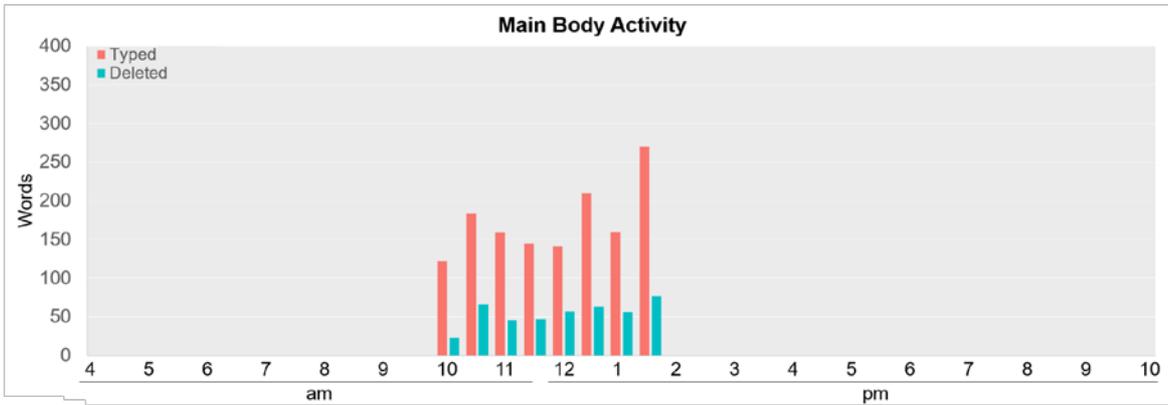
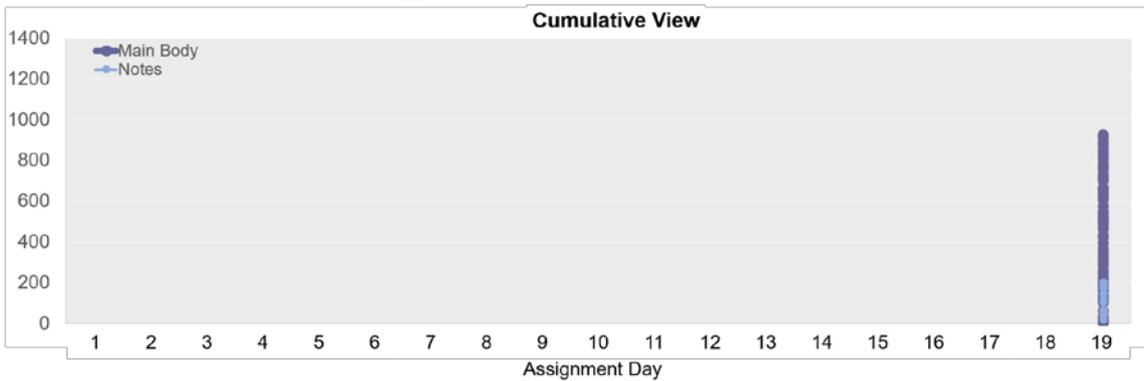
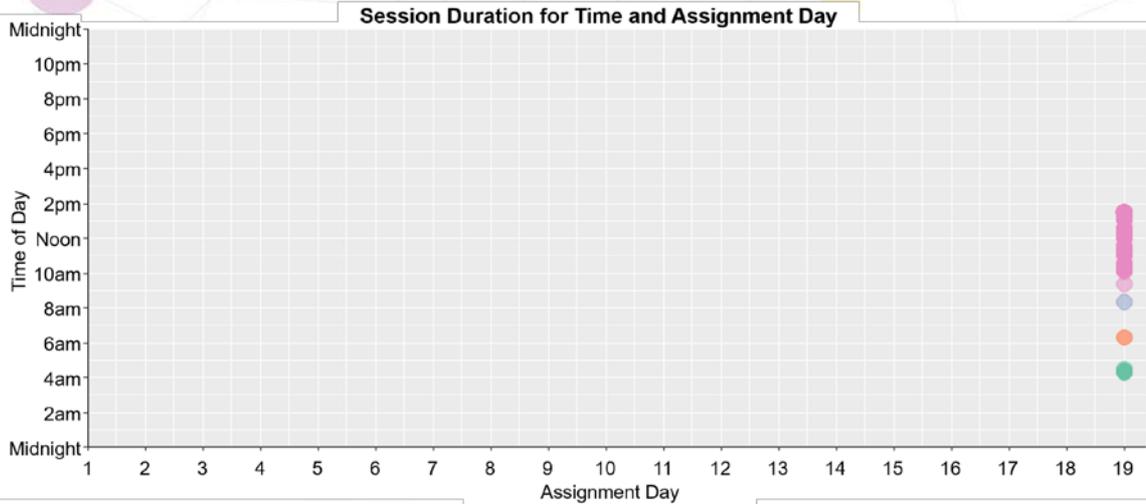


Figure 4: Case study 3: EF. Figure shows Session Duration (top) as a function of assignment day, and time of day throughout the assignment period (different colours indicates different sessions); Cumulative View (2nd from top) as a function of assignment day and number of words in the notes and main body; Main Body Activity (3rd from top) as a function of words typed and deleted over a single day; Notes Activity (bottom) as a function of words typed and deleted over a single day.

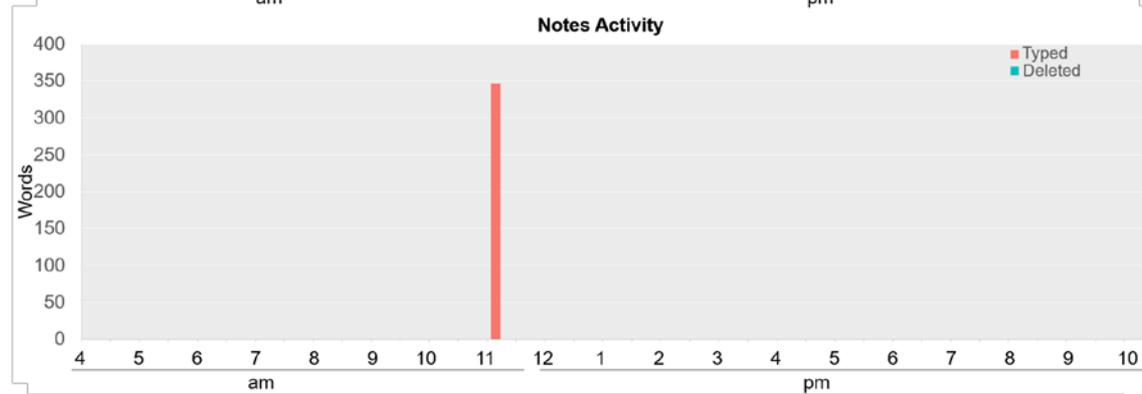
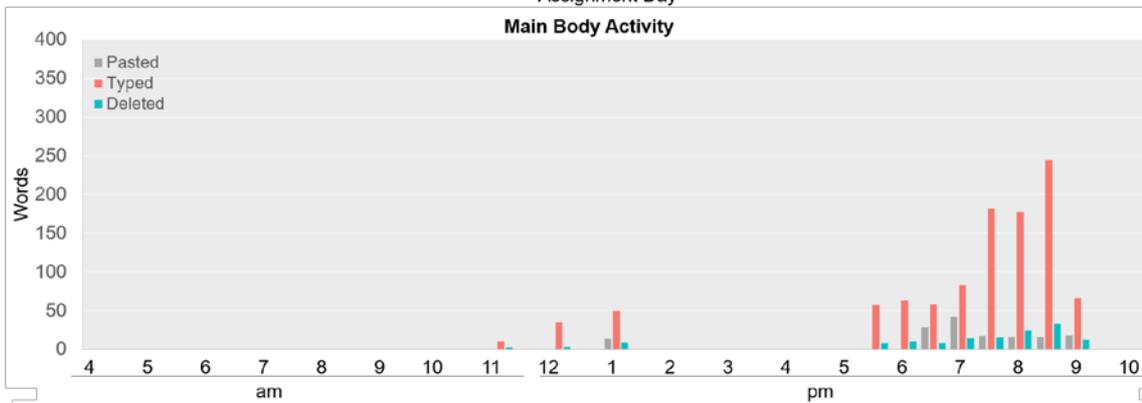
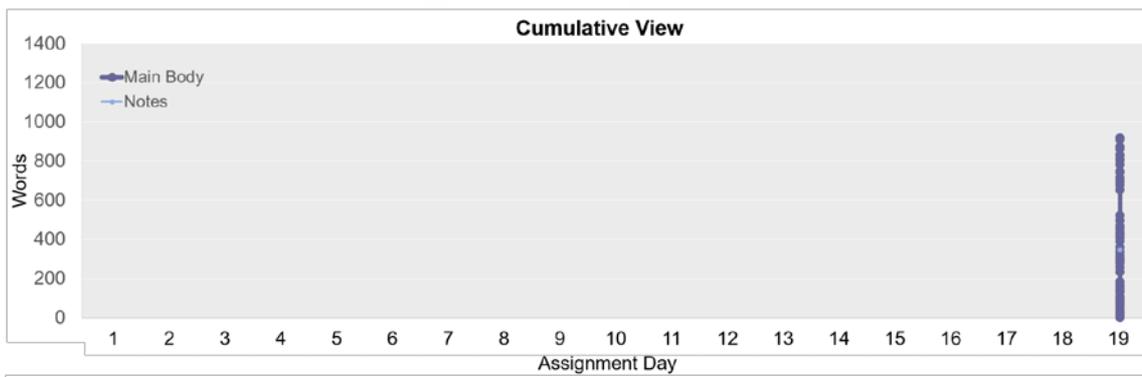
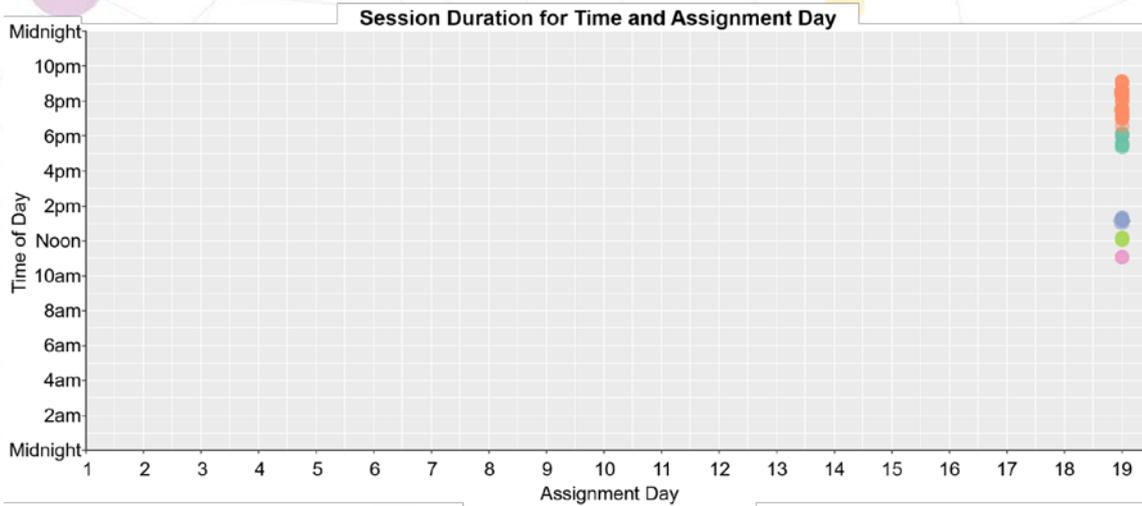


Figure 5: Case study 4: GH. Figure shows Session Duration (top) as a function of assignment day, and time of day throughout the assignment period (different colours indicates different sessions); Cumulative View (2nd from top) as a function of assignment day and number of words in the notes and main body; Main Body Activity (3rd from top) as a function of words typed and deleted over a single day; Notes Activity (bottom) as a function of words typed and deleted over a single day.

Case study 4 – GH (Figure 5)

GH completed the essay in one day (day 19), over five sessions. GH worked on the essay from 11am to 9pm on submission day, taking a break from 2 to 5pm. Sessions ranged from 4 minutes to 2 hours and 21 minutes. Active duration ranged between 4 minutes and 1 hour and 20 minutes, with a median of 11 minutes. GH first used the notes section, typing about 350 words. No deletion of words was recorded in the notes. This suggests GH was either writing down ideas or copying text from another source, and not revising. From 11am until 1pm, GH made small contributions to the main body section: typing 95, pasting 13, and deleting 14 words over three sessions, indicating GH was possibly switching between writing the essay and other activities. When returning to work on the essay at 5pm, GH had two consecutive sessions: between 5:40pm and 6:15pm (16 minutes active duration), and a second between 7pm and 9:15pm (1 hour and 10 minutes active duration). During those sessions, GH added 929 words, with few deleted and pasted words. The low deletion behaviour possibly suggests GH completed the essay in another platform and then typed it in Cadmus. GH's final grade was 78%.

Discussion

The case studies each managed their time across the assignment in different ways. AB and CD were more successful in managing their time to work on the assignment, dedicating more time overall to the assignment. On the other hand, EF and GH worked on the assignment only on the last day, suggesting poor planning and time management. Participants who started on the essay earlier deleted a far greater number of words, compared to the participants who began working on the assignment the day it was due, indicating these participants spent a greater amount of time revising their essay. The data from the cases also suggests that participants who were better at managing their time - indicated by them dedicating a greater number of days to work on their assignment - have more flexibility in their allocation of time to the assignment. Participants with fewer days to work on the assignment appear pressured by time to complete the essay in fewer, but longer sessions. Regardless of prior time spent on the essay, all case studies spent a long time working on the assignment on submission day. These findings suggest students who dedicate more time to their written assignments are able to spend more time revising their essays, which has been associated with improved writing quality (e.g., Hayes & Nash, 2013, Santangelo, Harris, & Graham, 2015). However, it is not clear the reasons that led students to use such different strategies to manage their time. It could be due to differences in time management skills, strategic approach, or perceived value of the assignment. Further research is needed to examine influences of these different time management patterns.

For the three cases who made use of the notes section, only two (CD and EF) made use of the notes section including adding and deleting words, indicating planning behaviour. It is possible that the other participants were also planning, but did so in the main text area of the Cadmus system, or indeed offline. Previous research has found that planning before writing an essay can reduce revising time (Kellogg, 1988), and guided planning activities may improve writing quality (Santangelo, Harris, & Graham, 2015). The current findings suggest that providing a notes section and instructing students to use them may be a good idea to promote the use of planning. It is important to note that researchers have been cautious to suggest for students to simply plan more: "If people who write well plan a lot, that does not imply that teaching people to plan a lot will help them to write well" (Hayes & Nash, 2013, p. 49). Therefore, it is advised for educators (and/or word-processing software) to provide guidance to students when promoting planning activities, such as generation and organisation of ideas, to see an impact on their writing quality (Santangelo, Harris, & Graham, 2015).

While the current study did not focus on investigating the relationship between students' writing strategies and performance, the selected case studies suggest a possible association between patterns in time management and writing strategies (e.g., revision) and grades. The first two case studies achieved a grade of 85 and 81, respectively, and the last two cases studies achieved a grade of 70 and 78, respectively. For these case studies, dedicating more days to work on the essay suggested to be related to the quality of their essay. Future studies may further examine the relationship between students' use of writing strategies and their performance.

Conclusion

This paper provides a novel approach into examining students' use of essay writing strategies in real-time and in a real-life setting. The four case studies illustrate students' similarities and differences on the use of self-regulation strategies when writing an essay in a real-life setting through the use of learning analytics. More specifically, we examined how participants manage their time and use planning and writing strategies while working on an essay over a 19-day period as part of their undergraduate course. Participants completed the essay over a different number of days, engaged in sessions of different durations, and at different times of the day. The participants used a variety of approaches to their writing. One participant typed many words within a single day and allowed significant time for editing. Another typed gradually, with many sessions over a number of days, and made a major contribution at the last minute. Two participants waited until the due date to complete the essay over a very long session. Overall, this study demonstrates the potential value of examining students'

writing process over long periods of time in real-life settings. While very preliminary, the findings suggest that these patterns may be related to students' performance. Most importantly, this study demonstrated that the word-processing software used was a useful tool for research purposes in essay writing. Further studies as part of this project will focus on examining students' use of specific features of Cadmus using a more in-depth temporal analysis to better understand students' writing processes and their relationship to the use of self-regulate learning strategies.

References

- Azevedo, R. (2014). Issues in dealing with sequential and temporal characteristics of self-and socially-regulated learning. *Metacognition and Learning*, 9(2), 217-228. doi: 10.1007/s11409-014-9123-1
- Baxter, P., & Jack, S. (2008). Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers. *The Qualitative Report*, 13(4), 544-556. Retrieved from <http://nsuworks.nova.edu/tqr/vol13/iss4/2>
- Biggs, J. (1988). Approaches to learning and to essay writing. In R. Schmeck (Ed.), *Learning strategies and learning styles* (pp. 185-228). New York: Plenum.
- 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 No. 14-03. Princeton, NJ: ETS.
- Eklundh, K. S., & Kollberg, P. (2013). A computer tool and framework for analysing online revisions. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 163 – 188). Hoboken: Taylor and Francis.
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College composition and communication*, 32(4), 365-387.
- Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64-71.
- Hayes & Nash (2013). On the nature of planning in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 29 – 56). Routledge.
- Kellogg, R. T. (1988). Attentional overload and writing performance: Effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(2), 355. doi:10.1037/0278-7393.14.2.355
- Mao, A., Kamar, E., & Horvitz, E. (2013). *Why stop now? predicting worker engagement in online crowdsourcing*. Paper presented at the First AAAI Conference on Human Computation and Crowdsourcing.
- Santangelo, T., Harris, K. R., & Graham, S. (2015). Self-Regulation and Writing. In C. MacArthur, S. Graham & J. Fitzgerald (Eds), *Handbook of Writing Research, Second Edition* (pp. 174 – 193). New York: Guilford Publications.
- Siemens, G., & Gašević, D. (2012). Special Issue on Learning and Knowledge Analytics. *Educational Technology & Society*, 15(3), 1–163.
- Stratman, J., & Hamp-Lyons, L. (1994). Reactivity in concurrent think-aloud protocols. Issues for research. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp. 89-112). [Written Communication Series 8]. Thousand Oaks, CA: Sage.
- Torrance, M., Thomas, G. V., & Robinson, E. J. (2000). Individual differences in undergraduate essay-writing strategies: A longitudinal study. *Higher Education*, 39(2), 181-200. doi:10.1023/A:1003990432398
- Van den Bergh, H., & Rijlaarsdam, G. (2013). The dynamics of composing: modelling writing process data. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 207 – 232). Hoboken: Taylor and Francis.
- Winne, P. H. (2010). Improving Measurements of Self-Regulated Learning. *Educational Psychologist*, 45(4), 267-276. doi:10.1080/00461520.2010.517150
- Zimmerman, B. J., & Risemberg, R. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary educational psychology*, 22(1), 73-101. doi: 10.1006/ceps.1997.0919

Contact author: Kelly Trezise,
trezisek@unimelb.edu.au.

Please cite as: 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* (pp. 411-421).

Note: All published papers are refereed, having undergone a double-blind peer-review process.