2018 ASEE Annual Conference & Exposition, June 24–27, 2018, Salt Palace Convention Center, Salt Lake City, UT

Paper ID #21710

Data Visualization for Time-Resolved Real-Time Engineering Writing Processes

Dr. Catherine G.P. Berdanier, Pennsylvania State University, University Park

Catherine G.P. Berdanier is an Assistant Professor in the Department of Mechanical and Nuclear Engineering at Pennsylvania State University. She earned her B.S. in Chemistry from The University of South Dakota, her M.S. in Aeronautical and Astronautical Engineering and Ph.D. in Engineering Education from Purdue University. Her research interests include graduate-level engineering education, including inter- and multidisciplinary graduate education, online engineering cognition and learning, and engineering communication.

Prof. Natascha Trellinger Buswell, University of California, Irvine

Natascha Trellinger Buswell is an assistant teaching professor in the department of mechanical and aerospace engineering at the University of California, Irvine. She received her B.S. in aerospace engineering from Syracuse University and her Ph.D. in engineering education from the School of Engineering Education at Purdue University. She is particularly interested in teaching conceptions and methods and graduate level engineering education.

Ms. Zixuan (Victoria) Zhao, Purdue University

Zixuan (Victoria) Zhao graduated from Penn State University in May 2017, where she earned her B.S in Mechanical Engineering. She has a passion for energy utilization and engineering education. She spent seven month working with Volvo Group Trucks and analyzed the energy consumption for heavy duty vehicles. Currently, she is a first year graduate student at Purdue University. Her research topics explore different materials to achieve efficient radiative cooling and better unitize the solar energy. She is also a course instructor for the undergraduate heat transfer lab at Purdue University.

Data Visualization for Time-Resolved Real-Time Engineering Writing Processes

Abstract

In this research paper, we present results of a new method for capturing and visualizing real-time data. Results presented represent nearly ten hours of real-time writing data from one graduate student applying for the NSF Graduate Research Fellowship Program. Though we show our analyses for only one participant, this methods paper demonstrates the use of novel data visualization tools to effectively "see" large qualitative data sets. Data was collected using screen capture techniques and coded using a validated coding schema facilitated with a dynamic "touch screen" coding interface to more easily code hours of authentic data. The visual representations of cognitive engineering writing patterns indicate several different aspects of "visible" cognitive writing processes, such as the iterative nature of the composing and knowledge-gathering parts of writing, and continual reference to the task materials that define the criteria upon which the written document will be evaluated. We anticipate broadening this study using these methods in order to develop heuristics for engineering academic writing, and to study the ways in which expert engineering writers overcome issues such as writer's block. The findings and representations of data as shown in this paper offer much to the engineering education research community in terms of method development and analysis of large quantities of time-resolved data representing authentic engineering communication skills.

1. Introduction

Despite the fact that national calls for engineering student knowledge, skills, and attributes emphasize the ability for engineers to be able to communicate verbally and in writing, little formal attention is paid to the theory-based teaching and learning of academic, disciplinary writing, especially at the graduate level. Even for engineering students pursuing careers in industry, many responsibilities in industry for graduate degree-holding engineers require strong written and verbal communication skills, and many engineers still publish research papers in conference and journal publications. Most graduate students are underprepared for their academic milestones (e.g., thesis, dissertation, papers). Students pursuing academic careers are especially underprepared for writing tasks such as grant writing, and those pursuing careers in industry are equally unprepared for the genres of writing required in the workplace. Prior work in engineering writing by the authors and others study writing in a "static" context: That is, final documents are analyzed in order to understand argumentation structure through a genre lens. Other work has sought to understand the ways in which writers may struggle with the writing process from an affective dimension. However, very little is known about the cognitive patterns of engineering writers writing in authentic disciplinary contexts.

In this paper, we present the methodological approach and data visualization of time-resolved writing data. For the purpose of this paper, we focus on one graduate student, Fred, as we describe the methods we used to visualize and represent his real-time screen-capture data. In the discussion, we examine what our approach makes visible about graduate student writing processes and describe how our work could potentially further the literature and understanding of engineering graduate student writing, engineering writing in general, and other large amounts of time-resolved data.

2. Literature Review

Engineering writing competencies at the undergraduate level are often emphasized in terms of Writing Across the Curriculum (WAC) and Writing in the Disciplines (WID) initiatives [1,2]. However, at the graduate level, academic engineering writing is rarely taught through formal mechanisms, instead relying heavily on research advisorship to teach academic writing skills [3–5]. While some "lucky" students naturally develop academic writing skills through a combination of strong advising, exposure with multiple genres of academic writing, and educational capital, [6] many graduate students are not introduced to the requirements of disciplinary academic writing until high-stakes deliverables such as master's theses, publications, or dissertations are underway. While the linkages between writing competencies and student success are loose, many pieces of research in both the graduate education literature and the writing literature point to the idea that the ability to write as a member of an academic discipline is a symbol of socialization,[7–10] academic literacy, [11–13] and that the inability to communicate effectively in writing can add years to the dissertation process or potentially result in attrition. [14]

The bodies of literature in engineering that involve engineering writing and education are generally quite separate from the bodies of literature in English composition and rhetoric that explore sociological and cognitive aspects of writing. Writing-focused literature in engineering education tends to be "interventionist," tending toward reporting on courses or best practices for encouraging writing habits in engineering courses. As examples, Leydens and Olds[15] proposed a disciplinary engineering grant-writing course based in genre theory, showing that authentic engineering writing tasks are most beneficial to teach graduate students the language patterns required for academic success. Similar to other initiatives trends across disciplines that capitalize on social learning theory through disciplinary writing groups, [6,16–18] the Dissertation Institute[19] seeks to coach engineering doctoral students from underrepresented groups through the dissertation writing process in order to increase completion statistics and decrease time wasted in the “ABD” (All but Dissertation) stage of writing. Most graduate-level engineering writing work fails to rely on the methods by which students learn to write, and do not capitalize on methods or findings from the English/writing research community, such as genre theory[20–23] which works to understand the linguistic ways in which sentence purposes interact to convey disciplinary messages.

While genre studies are useful in reveal the underlying patterns of disciplinary writing for engineering students,[20,21,24,25] one of the biggest gaps in the literature is that most of these studies investigate static, finalized, and "perfect" versions of writing, rather than demonstrating to students that real writing is messy, imperfect and iterative. Literature documents that oftentimes students struggle with perfectionism and writer's block,[26–29] among other barriers to writing, perhaps because understanding the "messy" parts of writing are part of the apprenticeship involved in gaining competencies in disciplinary communication. [30] While having final and polished documents can be useful to students as model texts [30] it is likely useful for students to be able to see other peoples' writing styles and the iterative nature of the writing process.[31] However, very few writing studies exist that examine the intermediate and iterative drafting, revising, and editing processes that take up the most time in the writing process.

The subdiscipline of cognitive writing research does use strategic methods to capture the ways in which people are thinking or holding multiple pieces of knowledge in the mind at the same time.[32,33] For example, research employing keystroke-logging methods have existed for several decades, working to understand writing fluency and the processes involved in writing.[34–36] Augmenting keystroke logging methods, eye-tracking methods are also used to understand the cognitive processes in writing [37,38] but the results rarely translate back to the applied disciplinary writing community, or to (in our case) graduate students or research advisors in order to augment writing pedagogy. Further, the fields of cognitive writing research are filled with limitations: some researchers rely on highly intrusive methods to understand writing cognition, such as think-aloud during the writing process itself. [34,39] While think-aloud approaches to understanding writing can help make visible what is going on cognitively in a writer's mind, it also potentially distracts the writer and does not allow for authentic writing processes to occur.

The ongoing research on which this paper is based seeks to combine the fields of engineering writing with cognitive writing research methods, in order to more effectively understand the disciplinary and sociological nature of writing at the graduate level. This unique niche is understudied in any of the related literature and serves an important role in more fully enabling graduate engineering students to develop the academic literacies to write as members of their disciplines. As established in previous work[40] , our team is developing methods to non-intrusively capture real-time writing data, and we have reported on the development of a coding schema by which screen-capture data can be easily analyzed. In the present methods paper, we demonstrate the next step in the process. Using a real-time coding mechanism, we demonstrate novel ways of data visualization that can convert hours of real-time coded writing data into useful visual representations that convey findings easily to multiple audiences.

3. Cognitive Writing Theory: The Hierarchical Process Theory of Writing

This ongoing study aligns with Hierarchical Process Models of Writing developed over the last several decades by Flower and Hayes[41–43] to describe cognitive writing processes. The Hierarchical Process Model proposes that writers hold multiple facets of writing in their head at the same time, "popping up" to higher levels throughout the writing process. A diagram of Flower and Hayes' Hierarchical Process Model is shown in Figure 1. Of note are the different levels of activities that pertain to both social and cognitive dimensions. One of the main tenets of this theory is that writers hold multiple processes at the same time, for example, composing text while also anticipating the audience or the venue to which a manuscript will be submitted. The development of this theory and model has extended over time, to which aspects of technology have been assumed into the model: Composing and revising on a computer is much different cognitively than composing and revising by pen-and-paper. Some of the facets of cognitive writing theory are visible—that is, they are easily tracked through visible outcomes manifested through writing (e.g., composition or revision), while some of the categories in the initial model might be invisible (such as planning or considering needs of the audience.) In using the Hierarchical Process Model as a guide to inform our studies, we made methodological choices about how screen-recorded real-time writing data should be coded based on the fact that we can only capture the visible cognitive processes, as described in the data analysis section and in prior work.[40]

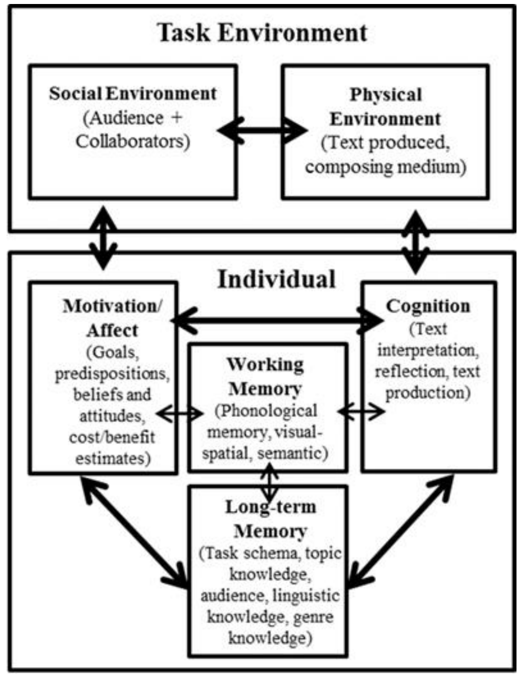

Figure 1. Hierarchical Process Model of Writing (Based on Flower and Hayes (1980, 1981, 2012)

4. Methods

A. Study Context and Participants

As part of the ongoing study, three graduate student participants were recruited to participate in the research process, which required them to record their computer screens using Camtasia screen capture software. To scope a relatively consistent, yet authentic writing task, the participants selected were all applying for the National Science Foundation Graduate Research Fellowship Program. In the award cycle of interest, the two deliverables included a three page "Personal, Relevant Background, and Future Goals Statement" and a two page "Graduate Research Plan Statement." For the purposes of this study, the research statements are particularly interesting as a miniature version of authentic disciplinary grant-writing. More information on the NSF GRFP and the criteria on which the applications are evaluated can be found via NSF.[44] Recruitment for this study began approximately three months before the NSF GRFP application was due. All participants are necessarily first or second year graduate students as per the application requirements for NSF GRFP, and are U.S. domestic students. We used pseudonyms to protect participant identities.

In the results and discussion sections of this paper, we focus on one participant, Fred, as he prepared his NSF GRFP application materials over ten hours of real-time screen-capture data. We elected to focus on one participant in order to highlight the methodological aspects of the data analysis and presentation. The data represented is ten hours worth of real-time writing data. In future work, we will compare the real-time maps across multiple participants.

B. Data Collection

After consenting to participate in the research project, all participants were given two licenses for Camtasia screen recording software; one that could be installed on a university computer, and one that could be installed on a personal laptop. The participants were instructed to start the screen capture software any time they were working on the written deliverables required for the NSF GRFP, regardless if they were actually "composing" new words. If the students needed to break in the middle of a writing session, they were able to pause the recording, or had the option to simply start a new recording when they returned. All video files were saved to a USB drive that was transferred to the primary researcher weekly during the data collection period. At this meeting when data was transferred from USB drive to the researcher's password protected external hard drive, the researcher also conducted a semistructured interview asking the participants to reflect on their writing process and any personnel resources they used during the week (e.g. if their advisor edited their documents, etc.). Hard copy "jottings" such as outlines or planning documents were also scanned and collected as part of the data set to support the screen capture data. These extra pieces of data are consistent with genetic artifact analysis methods typically used in linguistic and writing research.

Because the screen capture software simply runs in the background of a user's computer, it is entirely non-invasive, offering several methodological advantages compared to other real-time data collection methods (e.g. eye tracking setups that use goggles or headgear). As a result, all activities on the screen are also captured in addition to the screens solely devoted to the writing process: If a participant checks email, searches for literature, changes music, or instant messages a friend, all those activities are also recorded. Though the resulting data is messy, we argue that the "messiness" is actually demonstrative of an authentic writing process, which does not happen in a laboratory setting. In real life, the "writing" process of experts might require significant time searching for literature or checking manuscript/task requirements to comply with the evaluation criteria. The video data recorded offers a wealth of data to analyze. In our past work, we provide a literature-based commentary for the methods and methodological decisions that we made in developing an appropriate coding schema for the data.[40] These decisions were based on the epistemological decision-making frameworks from other disciplines who either use video data or other time-resolved data. The resulting protocol and coding mechanism is based on the premises of hierarchical cognitive writing theories, [42,43] and dictate that more than one class of activity can be done at a time, as the human mind is capable of handling multiple inputs during the writing process.

In this paper, we intend to present visualization methods for real-time writing data, with the hopes that other researchers can be inspired to re-envision non-traditional data in easily palatable ways. As such, we present the data and various visual interpretations of Fred's writing. Only his screen capture data is presented, not the associated jottings or interview data, to stay aligned with the goals of this paper. Fred's writing process over the course of several weeks consisted of several individual writing sessions that totaled nearly ten hours.

C. Data Analysis

We developed a codebook that is based on both *a priori* and emergent codes which was used to analyze the real-time screen-capture data. The full description of the codebook development has been published elsewhere (Blinded for review), but the main points will be described here. The codebook is based on the Flower and Hayes[43] model which includes cognitive and social process aspects of writing processes. In our codebook, we only included the "visible" aspects of these writing processes due to the nature of our data, namely screen-capture video data. For example, we did not include internal, and therefore, invisible, processes that happen in the mind, such as elements of motivation, working-memory, long-term memory, and reading. These facets of the engineering writing process cannot be captured by screen-capture video data of writing on a computer. These non-visible aspects of writing may be explored through interviews with research participants, but including that data is outside the scope of this paper.

Table 1: Codebook for Cognitive Writing Processes

LEVEL	DEFINITION OF LEVEL	SUBLEVEL	CODE	DEFINITION OF CODE
Control Level	Monitoring of process overall		Planning	Outlining, generating a list of questions, gathering requirements, sense-making
Process Level	Divided into internal and external processes involved in the process of writing text	Composing Processes	Composing	The visible act of composing text in a manuscript
		Revision Processes	Addition of New Text	Identification of an area needing detail and writing new text
			Editing	Local editing: Spelling, word choice, grammar, etc.
			Rewriting	Rewriting a sentence/paragraph from scratch
			Revising	Altering a sentence/paragraph to add value
			Reorganizing	Moving text around, including copying/pasting
			Deleting	Deleting text from the manuscript without replacing
		Task Environment Affecting the Process	Collaborators and Critics	Visible interaction with social environment of the task (incorporating collaborator comments, etc.)
			Technology	Interacting with the media and composing technology (highlighting, underlining, changing fonts and formatting, etc.)
			Task Requirements	Focus on the requirements of the manuscript
			Task Materials	Including source texts, notes, outlines, etc.: Looking at webpages and external sources for technical information
Resource Level	Includes internal memories and general-purpose processes that the processes at the other levels can call on		Attention	Lack of focus on specific task at hand
			No Apparent Activity	Nothing currently happening on the screen

We imported this codebook into a free, online real-time data capture platform called the Generalized Observation and Reflection Protocol (GORP),[45] which was developed by the Center for Educational Effectiveness at UC Davis and is intended for capturing classroom activities and instructor pedagogies in real time. The platform is completely modifiable, and therefore, we found it very convenient for this data set. In using the tool, we found it most convenient to “play” the real-time screen capture recordings on a desktop computer with a large monitor and have the GORP tool on a touch screen tablet, so that as the data played, the researchers could click-on/click-off of the activities that were or were not happening.

The GORP tool outputs time-series data into an excel spreadsheet, which we then uploaded into MATLAB[46] for data analysis and visualization. MATLAB was chosen for its ability to visualize data in a multitude of representations. By mapping the time at which each code was used over the entire writing process, we are able to see more detail about the full writing process. We made a few assumptions in order to analyze our data in this way: We assumed that 1) the data are mathematically continuous, such that they can be considered functions, 2) multiple codes could happen simultaneously, and 3) the lag time between events the video-data and when they were coded in the GORP tool were small enough to be ignored.

5. Results

Our research objective was to develop visual representations of time-resolved real-time writing data. We therefore present multiple representations of the same data set to help understand various facets of the writing process for one participant, Fred. First, the discrete data points (each code captured at a single point in time) were plotted on a three-dimensional scatter plot, with the writing codes along one axis, the time stamp (hours) on another, and the duration percentage (or the percentage of total time spent on that code at that time) on the third axis, using the MATLAB 3D plotting function, shown in Figure 2. This visualization does give some indication of the clustering effects of the codes at certain points in the writing process. It also gives and understanding that the writer spends a duration of time on some codes more than others. However, this is not the most intuitive way to present the ten-hour data set. To better interpret the data set, we then assumed our data were continuous, such that we could visualize the data in more sophisticated ways. Figure 3 was generated by vectorizing the 3D scattered data from excel and then using MATLAB griddata function to interpolate the surfaces. This helps us see the "peaks" of where Fred was paying his attention during the

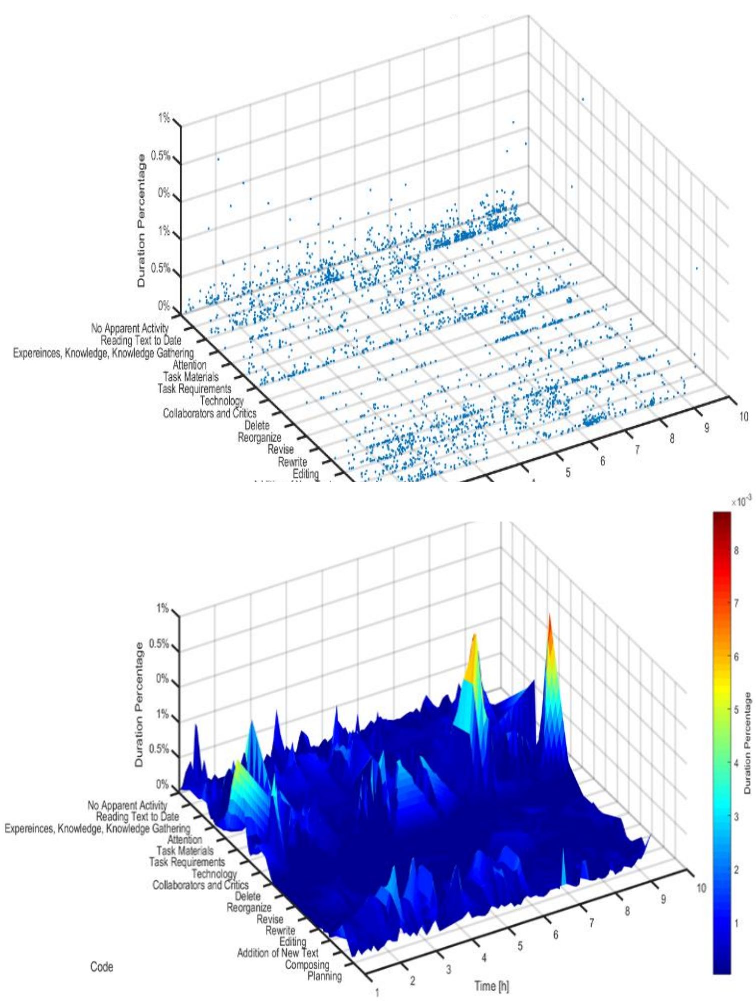

Figure 3: Vectorization of 3D plotted writing data

writing process. We see that Fred spent much of his time, especially near the end of the task, referring back to the "task materials" and "task requirements." This surface plot can be modified based on specific analytical goals. For example, we can “slice” the data (computationally, using MATLAB) along any one of the axes to show two-dimensional representations what is happening with one particular code or set of codes, or along a time stamp to see what behaviors occur near the beginning, middle, or end of a writing process. Similarly, the data can be envisioned from any direction, and re-colored as necessary to better show where Fred was directing his attention.

Most useful for this particular data set, we created a contour graph, or "heat-map," which allows us to see the data in another 2-dimentional representation in Figure 4. The color bar highlights where Fred's time was spent during the entire writing process. Blue represents short amounts of time, and red represents long amounts of time during the writing process. For example, in the beginning of the writing process, Fred took more time looking for task materials and task requirements. Looking at the light blue areas near the bottom from 2-6 hours, Fred spent most of the time composing, along with referring back to task requirement and utilizing technology. In the end, Fred focused on using technology for editing and formatting, as well as the relying heavily on the collaborators and critics code, as he incorporated last minute edits from his colleagues and research advisor before submitting his NSF GRFP application package.

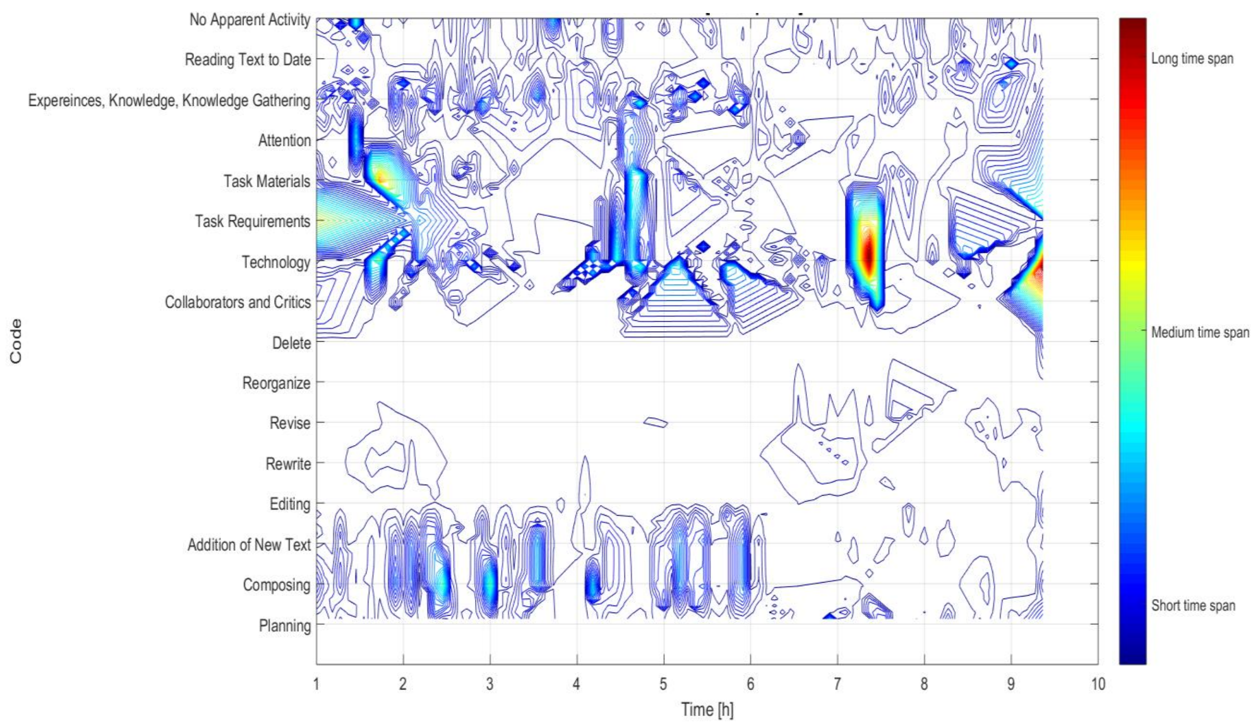

Figure 4. Contour plot of time, code and percentage of time that Fred spent on each task

Of particular note in this final representation is the cyclical nature of the planning and composing codes with the “higher order” codes where Fred was focusing on the task requirements and materials required to help him accomplish his task. The knowledge-gathering code—here, manifested through literature searches—was dispersed regularly throughout the first 60% of the writing process, and alternates with codes related to composing and the codes related to text modification (especially the addition of new text within previously written work). Futher, Fred re-visits the specific task requirements for the NSF GRFP several times throughout the process, spending time on them at certain points in the writing process. This indicates that Fred is actively checking that his documents are complying with the criteria by which the NSF GRFP is judged in order to make sure he has a strong chance of being favorably reviewed.

6. Discussion

The iterative and complex nature of writing is something that writing instructors often tell students, but rarely have the evidence to prove. However, in visualizing an entire writing process (here, ten hours' worth of data from the beginning to end of a writing task), we have shown the ability to "see" an entire writing process visually. Indeed, the cyclic, iterative nature of writing is intuitively shown through the heat-map visualization, and anecdotally, our students and colleagues have shown immediate understandings of the effort required for scholarly writing. We liken this process of "seeing" cognitive process data to that of Atman et al.[47] in their timeline representations of engineering design processes. Our representations help writers, educators, and researchers "see" writing processes over a long duration of time, and may be used to foster reflection in engineering writers at all levels.

These results demonstrate the usefulness of this method to visualize large quantities of data, in this case, related to a time-resolved visualization of a graduate student's writing process over nearly ten hours over multiple weeks. This method advances the writing literature, which typically studies time-resolved writing for brief, inauthentic tasks based on essay prompts. These methods cannot effectively capture the "messiness" of real writing, which does include using outside resources, time seemingly "inactive" during writing, and other inputs to the writing process, such as checking email, listening to music, or instant messaging with friends.

Ultimately, the goal of this larger process is not to demonstrate a "right" way to write. Correlating with the writing literature, each person has her or his own strategies and process by which they write, that correlates with their conceptions and attitudes toward the writing process. In this paper, we only show the results for one participant over ten hours of data; however, in the future we will expand this work to ultimately work towards writing heuristics by considering many writers' real-time data. In other words, we seek to understand general patterns of writing activities enacted by disciplinary engineering writers as they write authentic deliverables over long periods of time. Additionally, by presenting multiple persons' writing representations, we can definitively demonstrate that the writing process can be successful while looking different. Future work includes the development of real-time writing heuristics, showing how engineering writers at all levels (expert vs novice) might deal with writer's block, time management, or any other writing issues that may plague writers at all stages. We hope to translate these real-time writing results back into the engineering classroom in showing students the complexity of writing, and ultimately, that a one-time "cram" session is not sufficient to write for an academic or scholarly venue.

Limitations of this method include the simplification of actual human cognition: We cannot expect to see the actual internal planning process occurring in the brain of the writer. Instead, we focus on evidence of planning. For Fred, this meant scrolling up to his “brainstorming” section of his research statement document, for example, or writing bullet points or notes to himself to remind himself what he might need to talk about. Similarly, this method does not capture intent. When Fred left his manuscript in order to find additional literature on Google Scholar, we had no way to map his intentions or what sort of information he was looking for until he cited that paper in his manuscript. However, we can code those actions as “knowledge gathering” more broadly.

These limitations are overcome by the benefits of this method in being able to see a great deal of data on a two-dimensional plot. This non-intrusive method does not ask writers to think-aloud at the same time as they are writing, as prior studies have asked writers to do, thereby confounding the cognitive writing process. In addition to this data, we collected jottings, hard-copy planning documents or revisions, and also conducted interviews with the participant to capture some of the overarching rationales and attitudes about the writing process and writing strategies overall; however, that data has not yet been triangulated with the real-time writing data shown in these figures. In future work, we will bring in these additional data sources to help interpret the non-visible writing processes and work towards triangulating the processes we saw with the screen-capture data.

We expect that the results of this paper will offer readers techniques or ideas on how to use data visualization tools to begin to expand ways of visualizing qualitative data in palatable and creative ways. While we are using these tools to show cognitive writing processes, we hope to encourage other researchers who have unruly or non-traditional qualitative data that occurs as a function of time to begin to use alternative data analysis methods to best visualize the “story” in their data. While traditional methods for qualitative data analysis and presentation are quite good for some contexts, we recommend researchers use creative toolsets (and engineering data visualization toolsets, such as MATLAB!) to better visualize data.

7. Conclusions

In this paper, we presented the visual representations of one writer's real-time screen-captured data as he applied for the NSF GRFP. This data set included almost ten hours of time-resolved data that we coded for various writing elements. We represented this data in multiple ways that allow us to "see" the writing process as it occurs in an authentic task, over multiple writing sessions that occurred over several weeks. By producing visual representations of the writing process, in particular, a graduate student's writing process during an authentic task, we believe other graduate students and their advisors will have a better understanding of what "successful" writing can look like. We want to emphasize that writing is a messy process, one that does not happen linearly and is iterative, and believe that by sharing these results, more graduate students and faculty members who mentor those graduate students, will be able to reflect on writing processes. By better understanding graduate student writing behavior, educators can support graduate students through the critical and necessary process of writing up their research in disciplinary discourse. In addition to better understanding writing, we also feel that this work has large implications for other real-time and time-resolved data in educational settings.

References

1. Leydens, J. A. Sociotechnical communication in engineering: an exploration and unveiling of common myths. *Eng. Stud.* **4,** 1–9 (2012).

2. Paretti, M. C. & McNair, L. D. Introduction to the Special Issue on Communication in Engineering Curricula : Mapping the Landscape. *IEEE Trans. Prof. Commun.* **51,** 238–241 (2008).

3. Ross, P. M., Burgin, S., Aitchison, C. & Catterall, J. Research writing in the sciences: Liminal territory and high emotion. *J. Learn. Des.* **4,** 14–27 (2011).

4. Catterall, J., Ross, P., Aitchison, C. & Bergin, S. Pedagogical approaches that facilitate writing in postgraduate research candidature in science and technology. *J. Univ. Teach. …* **8,** 1–10 (2011).

5. Belcher, D. The apprenticeship approach to advanced academic literacy: Graduate students and their mentors. *English Specif. Purp.* **13,** 23–34 (1994).

6. Aitchison, C. Writing groups for doctoral education. *Stud. High. Educ.* **34,** 905–916 (2009).

7. Rose, M. & McClafferty, K. A. A Call for the Teaching of Writing in Graduate Education. *Educ. Res.* **30,** 27–33 (2001).

8. Bremner, S. Socialization and the acquisition of professional discourse: A case study in the PR industry. *Writ. Commun.* **29,** 7–32 (2011).

9. Austin, A. E. Preparing the Next Generation of Faculty: Graduate School as Socialization to the Academic Career. *J. Higher Educ.* **73,** 94–122 (2002).

10. Duff, P. A. Language socialization into academic discourse communities. *Annu. Rev. Appl. Linguist.* **30,** 169–192 (2010).

11. Lea, M. R. & Street, B. Writing as academic literacies: Understanding textual practices in higher education. in *Writing: Texts, Processes, and Practices* (eds. Candlin, C. N. & Hydlan, K.) 62–81 (Longman, 1999).

12. Johnson, E. M. Toward the building of a cross-disciplinary doctoral research and writing culture. *J. Univ. Teach. Learn. Pract.* **11,** 1–15 (2014).

13. Gustafsson, M. Academic literacies approaches for facilitating language for specific purposes. *Iberica* **22,** 101–122 (2011).

14. Gardner, S. K. Student and faculty attributions of attrition in high and low-completing doctoral programs in the United States. *High. Educ.* **58,** 97–112 (2009).

15. Leydens, J. & Olds, B. Publishing in Scientific and Engineering Contexts: A Course for Graduate Students. *IEEE Trans. Prof. Commun.* **50,** 45–56 (2007).

16. Carter-Veale, W. Y., Tull, R. G., Rutledge, J. C. & Joseph, L. N. The Dissertation House Model: Doctoral Student Experiences Coping and Writing in a Shared Knowledge Community. *CBE Life Sci. Educ.* **15,** 1–12 (2016).

17. Allison, D., Cooley, L. & Lewkowicz, J. Dissertation writing in action: The development of a disseration writing support program for ESL graduate research students. *English Specif. Purp.* **17,** 199–217 (1998).

18. Locke, L. A. & Boyle, M. Avoiding the A . B . D . Abyss : A Grounded Theory Study of a Dissertation-Focused Course for Doctoral Students in an Educational Leadership Program. *Qual. Rep.* **21,** 1574–1593 (2016).

19. Hasbun, I. M., Matusovich, H. M. & Adams, S. G. The dissertation Institute: Motivating doctoral engineering students toward degree completion. in *IEEE Frontiers in Education Conference***,** American Society for Engineering Education (ASEE) (2016).

20. Swales, J. *Genre analysis: English in academic and research settings*. (Cambridge University Press, 1990).

21. Luzón, M. J. Genre analysis in technical communication. *IEEE Trans. Prof. Commun.* **48,** 285–295 (2005).

22. Tardy, C. M. A Genre System View of the Funding of Academic Research. *Writ. Commun.* **20,** 7–36 (2003).

23. Wingate, U. Using academic literacies and genre-based models for academic writing instruction: A ‘literacy’ journey. *J. English Acad. Purp.* **11,** 26–37 (2012).

24. Kanoksilapatham, B. Rhetorical structure of biochemistry research articles. *English Specif. Purp.* **24,** 269–292 (2005).

25. Soler-Monreal, C. A move-step analysis of the concluding chapters in computer science PhD theses. *Ibérica Rev. la Asoc. Eur. Lenguas para Fines Específicos* **32,** 105–132 (2016).

26. Lonka, K., Chow, A., Keskinen, J., Sandstrom, N. & Pyhalto, K. How to measure PhD. students ’ conceptions of academic writing – and are they related to well-being? *J. Writ. Res.* **5,** 245–269 (2014).

27. Onwuegbuzie, A. & Collins, K. Writing apprehension and academic procrastination among graduate students. *Percept. Mot. Skills* **92,** 560–562 (2001).

28. Bloom, L. Z. Anxious writers in context: Graduate school and beyond. in *When a Writer Can’t Write: Studies in Writer’s Block and Other Composing-Process Problems* (ed. Rose, M.) 119–133 (The Guildford Press, 1985).

29. Bartholomae, D. Inventing the university. in *When a Writer Can’t Write: Studies in Writer’s Block and Other Composing-Process Problems* (ed. Rose, M.) 134–165 (The Guildford Press, 1985).

30. Lillis, T. & Scott, M. Defining academic literacies research: Issues of epistemology, ideology and strategy. *J. Appl. Linguist.* **4,** 4–32 (2008).

31. Alamargot, D. & Lebrave, J.-L. The study of professional writing. *Eur. Psychol.* **15,** 12–22 (2010).

32. McCutchen, D. A capacity theory of writing: Working memory in composition. *Educ.*

Psychol. Rev. **8,** 299–325 (1996).

33. Alamargot, D., Caporossi, G., Chesnet, D. & Ros, C. What makes a skilled writer? Working memory and audience awareness during text composition. *Learn. Individ. Differ.* **21,** 505–516 (2011).

34. Leijten, M. & Van Waes, L. Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Writ. Commun.* **30,** 358–392 (2013).

35. Johansson, R., Wengelin, Å., Johansson, V. & Holmqvist, K. Looking at the keyboard or the monitor: Relationship with text production processes. *Read. Writ.* **23,** 835–851 (2009).

36. Van Waes, L., Leijten, M., Wengelin, A. & Lingren, E. Logging tools to study digital writing processes. in *Past, Present, and Future Contributions of Cognitive Writing Research to Cognitive Psychology* (ed. Berninger, V.) (Psychology Press, 2012).

37. Alamargot, D. *et al.* Eye-tracking data during written recall. in *Past, Present, and Future Contributions of Cognitive Writing Research to Cognitive Psychology* (ed. Berninger, V.) (Psychology Press, 2012).

38. Van Waes, L., Leijten, M. & Quinlan, T. Reading during sentence composing and error correction: A multilevel analysis of the influences of task complexity. *Read. Writ.* **23,** 803–834 (2009).

39. Gill, A. & Nonnecke, B. Think aloud: effects and validity. in *Proceedings of the 30th ACM international …* 31–35 (2012).

40. Berdanier, C. G. P. & Trellinger, N. M. Development of a Method to Study Real-Time Engineering Writing Processes. in *IEEE Frontiers in Education* (IEEE, 2017).

41. Flower, L. S. & Hayes, J. R. The dynamics of composing: Making plans and juggling constraints. in *Cognitive Processes in Writing* (eds. Gregg, L. W. & Steinberg, E. R.) 31–50 (Lawrence Erlbaum Associates, Publishers, 1980).

42. Flower, L. & Hayes, J. A cognitive process theory of writing. *Coll. Compos. Commun.* **32,** 365–387 (1981).

43. Hayes, J. R. Modeling and Remodeling Writing. *Writ. Commun.* **29,** 369–388 (2012).

44. National Science Foundation. *Graduate Research Fellowship Program (GRFP) Program Solicitation*. (2014).

45. UC Davis Center for Educational Effectiveness. GORP Tool.

46. Mathworks, I. MATLAB.

47. Atman, C. J. *et al.* Engineering Design Processes: A Comparison of Students and Expert Practitioners. *J. Eng. Educ.* **96,** 359–379 (2007).