Development of a Method to Study Real-Time Engineering Writing Processes

Catherine G. P. Berdanier
Department of Mechanical and Nuclear Engineering
Pennsylvania State University
University Park, PA USA
cgb9@engr.psu.edu

Natascha M. Trellinger
School of Engineering Education
Purdue University
West Lafayette, IN USA
ntrellin@purdue.edu

***Abstract*— Engineering writing and communication is increasingly required of engineering undergraduate and graduate students; yet this call is rarely met through research-based methods. This research posits that deep investigation of the cognitive processes within engineering writing requires a novel method for data time-dependent analysis. However, no such method exists. This paper details our decision-making processes as we examine the available methods to analyze screen-capture video data of real time writing in an authentic engineering writing context. We detail the process of selecting the most appropriate approach and then provide a proof of concept for this method though analysis of screen capture video data of authentic engineering writing spanning several weeks. Complex issues in engineering education require creative interpretations of what can be used as "data," and similarly, creative and rigorous solutions for analysis. The results of this paper directly affect researchers doing real-time analysis of video data in all respects, and will also help other engineering education researchers to analyze non-traditional data.**

Keywords—method development; real-time data; engineering writing; methods and methodology

I. Introduction

In the crowded engineering curriculum, engineering writing and communication education competes for time with more technical classes, despite the constant call from industry and employers for graduates with better writing and communication skills. Often, the solution to this issue is to "outsource" technical writing electives to English departments taught by composition and rhetoric professionals. In general, most engineering students do not practice writing in a disciplinary context outside of laboratory reports in their undergraduate curriculum. For students in graduate school, often the first introduction to scholarly disciplinary engineering writing is in fellowship applications, master's theses, dissertations, and journal manuscripts written with their research teams. These are highly authentic tasks requiring exceptional writing skills, however, the skills required for writing effectively, efficiently, (and successfully) in these circumstances will usually not have been taught in the engineering curriculum.

In response to the work that has been done in English and Composition and Rhetoric in terms of writing process and the need to further investigate writing that is occurring in authentic engineering contexts, we propose that one area of research that is "ripe" for investigation is in the use of real-time data.

The present work delves into the methods development for this project and outlines our method development decision-making process heavily grounded in methodological literature. Although this paper presents our decision-making process for this analysis method development in the context of studying engineering writing in real time, we believe the method can be applied across contexts and disciplines. For example, we aim to help other researchers reconsider what can be used as "data" and model method development processes for others hoping to analyze non-traditional data.

II. Literature Review

A. Writing Research in Engineering

There is a lack of rigorous research in engineering writing and composition processes. A few researchers have forged interdisciplinary partnerships across the disciplines of English and Engineering [1-5] in order to more effectively study engineering writing in authentic settings. At the graduate level, a few genre-based studies have been conducted to analyze engineering fellowship proposals [6] and literature reviews [7] from a combined English and engineering research lens. The results of these efforts have served to continue important efforts in implementing Writing Across the Curriculum (WAC) and Writing in the Disciplines (WID) efforts that have been documented across disciplinary areas in the United States. One of the main limitations to these projects, while valuable, study the final product of writing (submitted grants, final literature reviews, etc.) and can not lend any insight into the writing process itself.

Cognitive writing researchers have proposed multiple complementary models to describe the writing process. For example, McCutcheon [8] advocates for a working memory theory of writing to explain how writers balance short term task-based memory (sentence structure, immediate editing, etc.) with long term working memory (cognizance of audience needs, task structure or goals, etc.) during the composing process. The works of Flower and Hayes through several decades [9-11] have developed a highly robust Cognitive

978-1-5090-5920-1/17/$31.00 ©2017 IEEE

Process Model of Writing. One of the advantages of the Cognitive Process Theory of Writing is that it notes that the process of writing is not linear, but rather is hierarchical and the interactions between the processes that a writer engages in and is not limited to a certain style of writer.

B. *Analysis of Real-Time Data*

We feel the future of writing research, and particularly in engineering, is to study writing in real time in order to investigate the ways in which writers conduct authentic tasks. The case of engineering is particularly intriguing since engineering curriculum rarely includes writing. Prior efforts in real-time writing include think-aloud protocols, keystroke analysis, eyetracking, and sceencapture methods [12-15] with the advent of the computer. One of the consistent issues with these method are that the method of collection often intrudes with the cognitive thought process. For example, thinking aloud while writing impedes the cognitive processes that authentic writing of disciplinary expertise requires, and has reported to be difficult in practice [16]. Similarly, bulky eyetracking hardware or external methods of videotaping a writer at work can be obtrusive and interrupt authentic processes.

Other work that has been done in real-time data analysis is behavioral work analyzing the functioning of teams [17], design thinking patterns [18-19], and in educational settings [20]. Usually, these tasks are filmed in order to understand some way of human behavior or their interaction with an artifact or task, rather than measuring cognitive processes. However, published studies from these settings can lend valuable insight into potential processes and methods for analyzing non-traditional real-time data.

C. *Purpose and Scope of the Present Research*

The purpose of this work was to develop a method by which to analyze real-time writing data that captured via screen-capture software. This fits into a broader long-term project that seeks to explore the writing patterns and processes of graduate engineering students as they write authentic documents for funding, as will be described in the Methods section further. This paper presents a description of the corpus of data used to develop the method, and explores the process by which the method was developed from best practices in literature from various disciplines. The results of this paper show analyzed data using our method, a discussion of the data as related to writing explicitly, and reflections and recommendations for other researchers working to analyze non-traditional or real-time data.

III. Description of the Corpus

A total of three students were "followed" as they applied for the NSF Graduate Research Fellowship Program fellowship, and many of hours screen-capture data were recorded for each student. In order to develop, refine, and validate our method, we chose just one student, Fred (pseudonym), to follow.

The data presented in this corpus is screen capture video recordings of Fred's writings: Before starting each writing session, he started Camtasia Studio software in order to capture everything that he was doing on the computer screen. The program runs in the "background" of the computer and is entirely non-intrusive except for having to start and stop the recording before and after a writing session. Draft versions of Fred's documents were also collected, and weekly interviews were conducted with the participant during his participation in the research project. For this research, we only consider the screen capture data. In total, there were ten writing sessions in the corpus, described in Table 1.

TABLE I. Corpus Description

Session	Date	Duration
1	10.04.15	00:10:18
2	10.12.15	00:46:06
3	10.13.15	01:04:43
4	10.17.15	01:20:49
5	10.17.15	00:56:34
6	10.18.15	01:50:02
7	10.18.15	00:29:46
8	10.19.15	00:51:32
9	10.25.15	01:05:15
10	10.26.15	00:56:36
	Total	**09:31:41**

As reflected by the timestamps, Fred varied in his writing session times on the computer. It is important to note that this approach did not capture any of the non-computer writing activities, such as discussing the documents with his advisor and peers, or planning on paper (however, these documents were collected during participant interviews). For the purposes of the present discussion, we will limit the method development to the analysis of video data only.

While the collected data is similar to traditional video data, such as a video of a classroom activity, it cannot be analyzed and understood in exactly the same way. In this case, the behavioral component to this research is the text production, cursor movement, and human-computer interaction is only visible on the computer screen, and thus, we can only consider what occurs on the screen for analysis. The cursor position and the generation, deletion, or revision of text, then, stands as a direct representation of the cognitive processes involved in writing.

IV. Methods Development

The following discussion outlines the methodological decisions we made in how to code and analyze non-traditional video data. Both methodological and methods-related decisions will be discussed, grounded in related literature. Our decisions were grounded in the experiences of researchers from a variety of fields who are working to understand non-traditional and real-time data; from educational researchers to human-computer interactions researchers to sociological geologists. While none of the contexts perfectly described a way to analyze our data, their best practices and methodological considerations helped to shape our priorities, as will be described.

A. *Methodological Considerations*

Within the literature on real-time and behavioral data analysis, methodological concerns are focused on accurately conveying the actions of the participants from a visual format into transcribed textual data, and then subsequently analyzing the data in light of the participants' motivations and intentions, rather than placing the researchers' lens on the data. Koschmann, Stahl & Zemmel [21] discuss the role of *ethnomethodological inquiry* in the analysis of video data, pulling from constructivist methodological traditions in that behavioral data is visible and situated. Similarly, Knoblauch and Tuma discuss the philosophical constraints within the interpretive methods required to understand video data: that "interpreting not only means describing and explaining non-verbal and/or verbal behavior but also determining the knowledge that one needs to understand what is going on in a situation and identifying the visible conduct that constitutes the situation (Heath, 1997)" [22, p. 419]. Many researchers warn against the temptation to regard video data as "neutral" and that the role of the researcher in the data analysis and "meaning making" phases is quite strong [23, 24]. Rather, the researchers need to have clear methods for the transcription of video data into written form (to be discussed in the next section), have strong mechanisms for interrater reliability and validity of coding [25], and be mutually dependent on methods and theory throughout the data analysis and interpretation phases.

Relationship with existing theory is equally important in the development of methods. Often this is phenomena-dependent, and will correlate strongly with the goals of the researcher. Some real-time data collection is conducted as a grounded theory ethnography, where all behaviors or activities are equally of interest. However, some researchers choose to have a more focused and theory-driven approach to research, even if they choose a constructivist paradigm from which to conduct their research. In these cases, the researchers will value certain behaviors, activities, or actions more highly than others. Neither approach is "right," as Langley [26] suggests. Rather, explicit determination of the goals of the research (in analyzing broadly and inductively without clear theoretical guidance versus closely examining the enaction of a micro-event through theory) is of utmost importance to the project.

Many of these suggestions are consistent with qualitative data analysis methods literature overall, but in light of non-traditional data to be analyzed and the development of a standard method, transparency in all methodological decisions and positionality of the researchers is important.

1) *In this Study: Positionality of the Researchers*

The authors on this paper are experts in studying engineering writing through several different lenses. Both are engineers by training, which lends them to interpreting disciplinary writing practices within the field of engineering. Within engineering education research, they have shared expertise in studying engineering writing. The first author on this study has extensive expertise in studying engineering writing and argumentation from a rhetorical and cognitive point of view, and the second author on the study has extensive experience in studying pedagogical applications for Writing Across the Curriculum initiatives in large engineering classes. Because of this positionality, we ground our decisions in engineering education and writing theory and in methods theory. In this work, we ground our research, methods, and methodological design firmly in research theory and writing theory.

2) *In this Study: Relationship with Theory*

Writing is both a cognitive and a social process: In composing any written work, cognitive scientists and composition and rhetoric researchers posit that there are several competing cognitive tasks juggled at any point in time [8-16]. In this study, we take a constructivist and interpretivist ethnomethodological approach to our data analysis and method development process. In such, for our particular task of studying real-time writing data from screen-capture video of the writing process, we assume that the enacted processes may be mapped onto various existing theories for cognitive writing processes. As such, we understand that many of the cognitive processes of writing are invisible. However, we assume that the composing processes and behaviors that are captured on the screen are representations of the cognitive writing processes themselves that are going on in the participants' mind. Sometimes the cognitive processes are enacted, such as when the participant is making a list of topics that should be covered, outlining, and making comments related to the task at hand. These all are evidence of a planning process going on in the mind. Similarly, if the participant breaks from the writing task itself in order to look up technical information or literature related to a research proposal, we can say that the participant is gathering information related to the task. However, if the participant is inactive for a period of time, we cannot for certainty say if he is reading his work, planning the next thought, or if he is actually at the computer at all (perhaps talking to a colleague, eating a snack, or texting). This fundamental assumption checks us as the researchers from over-inferring cognitive processes from what is enacted on the computer screen through the writing data.

Our priorities are reflected in our decisions: Based on literature, we felt it necessary to record the pauses that the writer took, recording them as "apparent inactivity" rather than drawing conclusions that he was not engaging in writing just because there was no text being produced. Through theory, planning and information-gathering is part of the writing process as much as composing is. We also decided that immediate correction of typos while composing did not constitute deletion or spelling correction as "revision" when it was embedded within a span of composition, because the global focus of the project is to study overarching patterns. Deletion of an entire sentence or cut-and-pasting a sentence from one paragraph to another, however, did constitute revision and engagement with technology. The other overarching value represented by our coding and other decisions was the way in which we recorded the "engagement with technology" code. Since the writing process was recorded using a computer, we could have felt compelled to record that the participant was engaging with technology at all times; however, this didn't serve to illuminate any parts of cognitive writing theory for us.

However, engaging with technology as the participant used the "strikethrough" function rather than simply deleting text, or using the autocorrect spelling function to correct grammar/spelling errors was decided to be interesting, even if the immediate goal was to study writing processes.

B. Literature-based Best Practices for Non-Traditional and Real-Time Data Sets

In this section, we highlight the most important considerations from literature in determining a standard method to analyze non-traditional and real-time data. Across methods literature, three interrelated considerations are generally cited as most important: (1) Development of the coding mechanism and codebook; (2) The consideration of unit of analysis (time scale); and (3) Development of a transcription process. Each of the decisions on these components should also align with the methodological paradigms with which the researcher determined previously. Unfortunately, there is also not an "order" in which to make these decisions: They are inextricably linked with each other and as such, the method development process will be recursive in nature.

The coding mechanism for the data set will be a crucial decision for the researchers, again, entirely dependent on the methodological paradigms of the researchers and the goals of the research project. Research approaches taking a grounded theory approach will follow the codebook development techniques of the paradigm with which they most closely align (positivist, post-positivist, or constructivist). Alternatively, researchers can use existing theory to provide a basis for coding their data. One caveat to using existing frameworks as coding mechanisms is that many frameworks have been developed through traditional qualitative research methods with data that is not developing in real-time, such as the analysis of interview protocols; document analysis techniques, etc. These codes might manifest in different ways based on what can be "seen" in a non-traditional data set, and likely the frameworks-as-codebooks will need to be revised based on the assumptions made by the researchers regarding units of interest. Further discussion on the development of qualitative coding techniques or quantification of qualitative data will not be discussed here, but the authors refer the readers to foundational methods texts.

In real time data analysis, there are a variety of ways to define the unit of analysis; and these ways have to do with the time scale of interest and the transcription mechanism (to be discussed further). Any decision, though, represents a value system of the researchers, and as such should be carefully considered. One way of capturing writing data is to take interval-based data, i.e. record what has happened in a set interval of time. For fast-paced data, this might be every five or ten seconds; for slower data, this could be minute to minute. This results in an even distribution of data collection that maps easily to a timeline of activities. However, depending on the phenomena of interest to the researchers, some behaviors, actions, or activities might mean very little, while others might be quite significant. (This is an area where the researchers' relationship to theory impacts the interpretation process heavily.) As an alternative to the interval-based data recording methods, some researchers gravitate toward "critical event" analyses, where the starting time and ending time of some behavior, activity, or event of interest is recorded and all other data is ignored. Critical event unit-of-analysis decisions are usually framed in light of a particular theory governing the phenomena of interest; however, even broad inductive approaches might use this method. For example, if the topic of interest is verbal communications in team settings, then the verbal exchanges are of interest, where the nonverbal behavioral data will not be analyzed. Within the verbal exchanges, though, the time scale of analysis might be related to a certain type of verbal exchange based on theory, or could be interval-based. These decisions are often both theoretical/methodological and pragmatic in nature (i.e. what works best for the data set in general) especially for non-traditional or real-time data. Positionality of the researcher can also taint this process; most sources recommend, "social coding" by more than one researcher at the same time in order to call attention to these biases [23, 27,28].

A final decision in the method development is how the data will be transcribed. *Transcription* is the process by which the data from a visual and/or auditory or non-written format is converted into a written data set that can then be analyzed. Many researchers offer advice in this area: See [22, 24 29-32] for further discussion and examples of transcription decisions. Based on the time-step and unit of analysis decision, transcription of the video data could manifest in one of several ways. One way would be that the researchers, at every unit or time of interest, would "report out" on the actions that had taken place over that unit of interest. After completing all of that transcription for the data set, the researchers would go back and verify that the transcription was accurately describing the activities. In this transcription mechanism, care must be taken by the researchers to not infer any rationale or motive to the actions or events. Indeed, as in ethnographic methods, it is a call to "make the familiar stange" to best capture phenomena [33]. After this transcription process, the transcription then becomes the data set, which is coded through either grounded theory or emergent/open coding mechanisms. Bezemer and Mavers [24] further address the theoretical considerations in the transcribing process.

Alternatively, some researchers, such as Snell [32] advocate for directly coding the data in real time, using commercially available qualitative coding software. Of course, this method works best after a relevant coding schema has been developed for a phenomena of interest, and the researchers are comfortable with the codebook and have already established strong interrater reliability. In the case of Snell's publication, the video data is still the primary data set (not a transcription of the video data) and the coding is done directly from the video data and not the transcription. Reliability and validity are of course of interest to both decisions, and the decision as to whether the transcription is accurate enough to become a primary data set will be up to the researchers and the data set of interest. This conversation also hints at the "spectrum" of transcription processes, and how transcription may or may not evolve over the course of a research project.

1) In this study: Development of a Codebook

With the understanding that all these decisions are related and iterative in nature, we found it most effective to start with multiple ideas for the coding mechanism, and then deduced the units of analysis and transcription process based on how it worked with our particular data set. Initially, we watched approximately 20 minutes of video data together and then proposed a variety of frameworks from literature that we thought might be a good basis for *a priori* codes for the data we had seen. A second viewing of the same interval of data quickly brought to light frameworks that were good in theory but would not be able to capture the phenomena of interest in our study. We arrived at an adapted version of the Flower & Hayes Cognitive Process Model of writing. One of the drawbacks of the initial theoretical framework was the fact that there are many parts of the model that are "unseen"—that is, we could not positively match a thought process with the activity happening in the writing process. For example, the Flower and Hayes model [11] deeply investigates the language processing patterns that happen in the mind in order to translate concepts in the mind onto written words on a manuscript or computer: We reframed the model for our purposes to only include the "visible" parts of the model that could be manifested in the screen capture writing data. For example, there would be no possible way to capture if the participant was accessing his short or long term memory functions in the screen-capture data.

Throughout the transcription and coding process, (which will be further discussed), as activities, behaviors, and events came up that did not seem to fit within the categories in the theory, sub-categories were generated in order to fit the data and describe the enacted writing strategies better. Table 1 shows the finalized codebook.

TABLE II. CODEBOOK FOR COGNITIVE WRITING PROCESSES

Level	Definition of Level	Sublevel	Code	Definition
Control Level	Monitoring of process overall		**Planning (Sensemaking)**	Outlining, setting a list of questions, gathering requirements
Process Level	Divided into internal and external processes involved in the process of actually writing text	**Composing Processes**	**Composing**	The visible act of composing words in a manuscript
		Revision Processes	**Addition of New Text**	Identification of an area needing detail
			Editing	Local editing: Spelling, word choice, grammar, etc.
			Rewrite	Re-write a sentence/paragraph from scratch
			Revise	Alter the sentence to add value without rewriting the sentence
			Re-organize	Moving text around, includes copy-pasting
			Delete	Deleting text from the document without replacing
		Task Environment (External environment)	**Collaborators & Critics**	Visibly interact with social environment of the task (incorporate collaborator comments, etc.)
			Technology	Interacting with the media and composing technology (highlight, underline, fonts and formatting, etc.)
			Task Requirements	Focus on the requirements of the proposal
			Task Materials	Includes source texts, notes, outlines, etc.: Webpages and looking at 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 the specific task at hand
			No apparent activity	

As compared to Hayes [11] and related works, the invisible cognitive processes were merged together to reflect visible processes. In addition, we added a code for "no visible processes" understanding that while this could mean a lack of attention, it could also mean that the participant is thinking or jotting notes in a notebook. In addition to adjusting our codebook to only include observable activities, we recognized the need to capture context-specific elements as well as computer elements. In the case of this data, Fred is writing for a very specific purpose and is trying to convince his audience to fund his proposal. Our codebook is adaptable to this use of resources through the "Task requirements" code, and as he searches the internet for technical literature to better support his claims, we capture this knowledge-searching process as "task materials." The fact that this writing takes place on a computer adds an additional dimension to this data, and another one that we feel is necessary to capture (and has not been explored in writing literature well to-date). For example, we glimpse the song name pop up in the corner of the screen from the music he is listening to, or a message from his significant other pop up in a chat window. These little attributes reveal realities of the writing process that are more indirect and a result of our digital world, and are captured in the "technology" code.

2) In this study: Unit-of-Analysis and Time Scale Decisions

We deliberated whether to use an interval-based unit of analysis or a schema based on critical events. After trying multiple ways, we decided that it was important to us to (a) know what was going on in the data set at every point in time, but (b) be able to pinpoint exactly when different codes were used and for what precise duration of time. An interval based method would likely miss very short events, such as copy-pasting a portion of the document from one section into another; arranging argumentation structures, or moving quickly from the planning portion of the document back into the composing part of the document (for this participant). With an interval-based method, it would "look" like this code had been applied for the entire 10 or 20 seconds, when in fact, it was nearly instantaneous. Similarly, if the activity would happen at the junction between two timeslots, it could "look" even longer, or could get missed entirely.

3) In this study: Transcription Process Decisions

The transcription mechanism that we developed for this data set was built on the logistical considerations of the researchers' proximity to each other and the complexity of the data set. Originally, we had thought that our coding mechanism was simple enough we could code in real-time without transcribing; however, we quickly found that there were instances of activities that we wanted to know were going on, even if they didn't map explicitly into our coding schema because they weren't of interest to *this* particular study. One example of this is the time stamps on the videos: We were interested in the fact that our participant pauses the recording software for small bits of time—potentially to take breaks or check email (speculation on our part)—but then returns to writing later. We hope perhaps in the future to look at the time stamps to optimize students' ideal writing durations for maximum efficiency—however, this is not of particular use to us in the foundational writing process study.

For our purposes, we watched the video data together and did all data watching, transcription, and coding with each other. Though we were not together, we shared screens via Skype so we could talk to each other while watching the video data. We had clear definitions of roles within the project. One of the researchers is in charge of the video and a verbal transcription of the video data: Naming time points and narrating the activity (often coding at the same time) and stopping/replaying video data when necessary. At the same time, the other researcher—having watched the same bit of data—was responsible for writing down the narration in words in an Excel sheet and proposing code at the same time. The second researcher was also watching the video data for critical events related to the writing process that perhaps were missed by the first researcher. Both researchers are explicitly allowed to ask to stop the video, re-watch any section, or to discuss the articulation of the transcription or the coding of the transcription at any point in time. In this way, the "social viewing" mechanism advocated by Derry et al [28] helped to maintain the validity of the data. Portions of video that had multiple events happening in quick succession were watched as many times as necessary.

Though some sources in literature took a hard stance in terms of transcribing the data set entirely before coding it, we rejected the idea that then the video data would not be consulted in the coding process if we did this. Instead, we coded at the same time as transcribing. For an event of interest, we would stop the video, record the time, researcher #1 would narrate what had happened while researcher #2 typed it into the Excel sheet, adding any additional observations, and then both researchers would agree on the code or codes that the event represented. Though it seems like multiple steps, this turned into a very natural and streamlined process and often, it could be done in real-time along with the video.

Any pauses in activity greater than 10 seconds was coded as "no apparent activity," and often these codes fell within periods of composing, of planning, or of consulting the task requirements or materials (where the participant may have been reading the screen.) These long sections were easy to code in real time, whereas other periods were characterized by flurries of activities. An example of a "flurry" might be where the participant would move quickly from copy-pasting a criterion from the NSF GRFP website into the manuscript document (codes: technology; task requirements); scroll up to the planning outline and then back down to the body of the manuscript (planning; technology); compose a sentence or two (composing); tweak the wording of the sentence to be more concise (revise); delete a previous part of a sentence to merge with the new sentence (delete; revise); then pause for a bit of time (no apparent activity); and then continue composing (composing). These "flurries" were highly complex, and often took several viewings, discussion, re-windings, and re-viewings of material in order to capture the transcriptions and timestamps accurately.

V. Results and Validation of the Method

The results of this method development paper are presented here through the results of a small sample of our data. In this section, we present examples of the transcription and coding sheets used, as well as a representation of the data after analysis, grounded in writing theory. Through the description of the results presented in relationship to theory, we show high levels of construct validity. The reliability of the method is established through the expertise of the researchers on this study in investigating engineering writing and the transparency with which the methods and methodological considerations were developed.

A. Transcription, Coding, and Representation of Data in Alternative Forms

As described above we decided to indicate the start and end of each "critical incident" and include a description of the activity as our transcription. Table 3 shows the description of activities for a short excerpt of real-time writing data. Figure 1 is presented complimentary to Table 3, including the codes for the same excerpt of time. Here we see a "flurry" in action, where our participant transitions between many different activities in a short amount of time. It is particularly interesting to see that all these components constitute part of the writing process. It is important to note here that even though the first time stamp is at 12 minutes and 28 seconds, this is in actuality the third video file and in total the participant had been working on the writing task for several hours.

TABLE III. EXAMPLE OF TRANSCRIPTION OF SCREEN-CAPTURE DATA

TIME STAMP	DESCRIPTION OF ACTIVITY
12:28	composing
12:40	revising
13:00	crossing something out
13:13	deleting
13:14	paused
13:32	back on the internet
13:40	got an IM, in a chat conversation
13:50	back to composing
14:18	pause
15:15	back to planning section
15:17	paused
17:00	copied and pasted a planning sentence down in the composing section
17:20	paused
18:05	switched to IM chat
18:10	back to document composing
18:40	pause
19:40	back to composing
19:50	pause
20:08	technology chat
20:09	composing
20:44	pause

The resulting transcribed data was then coded using the a priori writing framework. This was also done in the Excel coding spreadsheet, marking the cell that corresponded to the appropriate code. In practice, our timestamps in our codebook have codes along the top row as headers and the timestamps are represented top to bottom. However, if this orientation is flipped such that time as an independent variable is presented horizontally, a story of this "flurry" begins to manifest. This strategy for data presentation is shown in Figure 1, where we also represented the codes from each of the "levels" of cognitive writing theory in different colors (e.g. Control level=Purple; Composing level=green; Revision=Yellow; Task Environment=Red; and Resources level= Orange). This is one way to "see" the data represented over time.

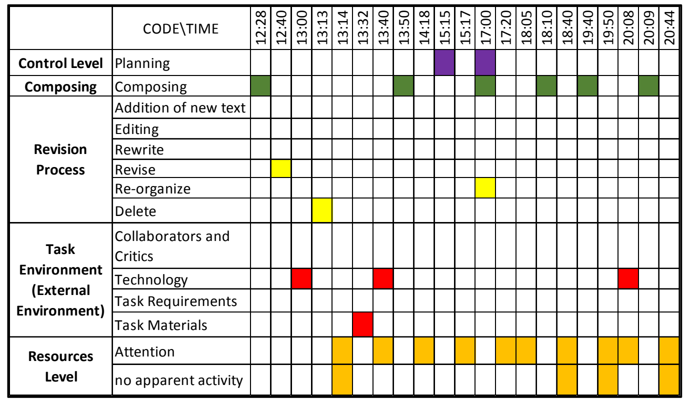

FIGURE I. CODING OF TRANSCRIBED DATA

Though this is a helpful representation of the data, this representation does not tell the entire "story" of the data, remembering that the intervals at which critical events are recorded and coded do not represent equal intervals of time. This has the effect of visually masking the effect of pauses in the writer's actual processes, nor does it reflect accurate the durations of time in which the writer was composing or revising.

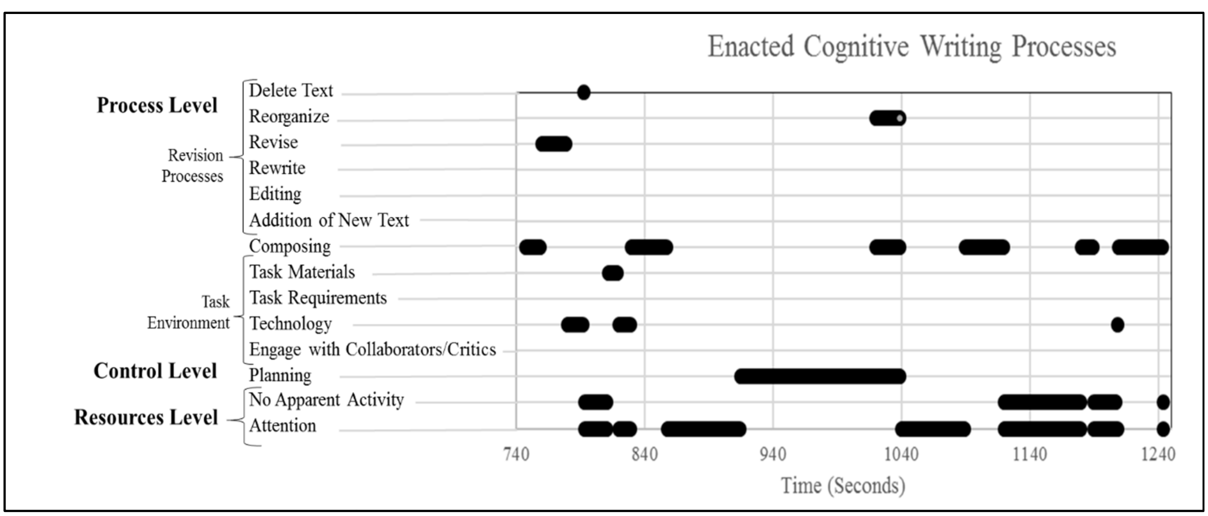

FIGURE II. REPRESENTATION OF ENACTED COGNITIVE WRITING PROCESSES

Rather, since time is a continuous independent variable, we can map the writer's activities on a line graph in order to visualize a more realistic writing process as shown in Figure 2. In this plot, the timestamps have been converted to seconds (rather than minute:second format) in order to more easily visualize the progression of time from the same window of time.

Of course, this is one 8-minute excerpt from the 10 hours of data from this participant, but it does represent much more about the way in which Fred engages with the writing process, especially how much time he spends composing in relationship to planning and not engaging in any apparent activity. One thing to note is that even though all of these things are part of the authentic writing process, the composing itself is by no means "fluent." Since this excerpt was taken from early in the set of data (approximately 2.5 hours from the beginning of the task), fewer codes representing the revision processes or engagement with collaborators/critics are shown.

VI. DISCUSSION AND APPLICATION TO OTHER RESEARCH CONTEXTS

A. Interpretation of the Presented Results through Cognitive Writing Theory

This format much more accurately represents the writing process. Of particular interest is the duration of time that the participant spends in "no apparent" action. While prior studies sometimes measure writing skill through fluidity (as measured by continuous writing/typing or lack of pauses in keystrokes [12, 34], this data indicates that the long pauses are often followed by information gathering (interacting with the task environment or technology), reviewing the task guidelines for insight into the criteria for success, or by a long period of composing. This indicates that the "break" is likely not due to inattention or "bad writing skill" but is part of a reflective practice from the writer that spurs another bout of text production or idea generation. In actuality, this participant's final documents and application package was strong enough to win him an NSF GRFP Honorable Mention, which is a significant achievement. Therefore, these interesting patches of inactivity and revision do not at all reflect the quality of the final product, only an unveiling of the cognitive processes involved in authentic writing tasks.

The findings from this research offer much to engineering communications researchers and technical writing instructors. First, this method is a way of "seeing" the thought process of writers, and it is highly likely that when we analyze the other two participants from whom real-time writing data was collected, the process might look radically different. Second, experienced writers who are skilled in writing within their discipline may have patterns of thought established that lend themselves to different patterns of composing, revision, and information gathering. However, we expect that the development of a method to analyze real-time writing data will unveil the fact that the revision and planning process happens throughout the writing process, not just at the end. In addition, we expect that this research will unveil the "normalcy" of taking many breaks during and between writing sessions, leading to a more holistic conversation about what "good" writing processes look like in practice.

B. Recommendations for Other Researchers Attempting Non-Traditional Data Analysis

As a helpful resource to other researchers considering the collection and analysis of non-traditional or real-time data, we present a list of reflections based on our experience that can serve as recommendations.

- Before making any decisions with regard to coding mechanism or any other analysis decisions, we highly recommend that other researchers think divergently about other fields that might be able to lend insight to data analysis.
- Even after exhausting methods literature, there may not be "answers" to how to best capture or transcribe the data.. Therefore, it is important to take both an inductive and a deductive approach to method development.
- Team structure and researchers will be different—team size, division of work is as important to the method as the codebook.
- It is important to be well-grounded in literature and theory such that research teams are able to articulate the possibilities and align them with epistemological and ontological values about what "matters" in the analysis process.

VII. CONCLUSION

In conclusion, this paper presents the process by which a method to analyze real-time writing of engineering graduate students was developed. The broader purpose of this article was to illuminate the decisions that must be made in interpreting non-traditional qualitative data, justified by best practices in literature and grounded in our own experiences. Results indicate that this method will continue to unveil interesting results for real-time cognitive writing. We also present reflections and recommendations for other researchers developing methods to analyze non-traditional data.

ACKNOWLEDGMENT

We acknowledge the participants that faithfully recorded their own data as part of this research project, and the research advisors and mentors that provided feedback on the preliminary research ideas related to engineering writing.

REFERENCES

[1] Berdanier, C. G., McCall, M., & Mike, G. (2016, October). A degree is not enough: Promoting engineering identity development and professional planning through the teaching of engineering résumé writing. In Frontiers in Education Conference (FIE), 2016 IEEE (pp. 1-5). IEEE.

[2] Berdanier, C. G., McCall, M., & Mike, G. (2016, October). Résumés in the development of undergraduate engineering identity: A genre analysis with teaching implications. In Professional Communication Conference (IPCC), 2016 IEEE International (pp. 1-9). IEEE.

[3] Paretti, M., McNair, L., Belanger, K., & George, D. (2009). Reformist possibilities? Exploring writing program cross-campus partnerships. WPA: Writing Program Administration, 33(1-2), 74-113.

[4] Leydens, J. A. (2008). Novice and insider perspectives on academic and workplace writing: Toward a continuum of rhetorical awareness. IEEE Transactions on Professional Communication, 51(3), 242-263.

[5] Leydens, J. A., & Olds, B. M. (2007). Publishing in scientific and engineering contexts: A course for graduate students tutorial. IEEE transactions on professional communication, 50(1), 45-56.

[6] Berdanier, C. G. (2016). Learning the language of academic engineering: Sociocognitive writing in graduate students (Doctoral dissertation, Purdue University).

[7] Berdanier, C. G.P. and Lenart, J. (2017). A Genre Analysis of Graduate Student Literature Reviews in Engineering: Toward Understanding Patterns of Disciplinary Argumentation. Accepted for In Professional Communication Conference (IPCC), 2017 IEEE International. IEEE.

[8] McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. Educational psychologist, 35(1), 13-23.

[9] Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. College composition and communication, 32(4), 365-387.

[10] Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. I LW Gregg & ER Steinberg (Eds.), Cognitive Processes in Writing (pp. 3-30).

[11] Hayes, J. R. (2012). Modeling and remodeling writing. Written communication, 29(3), 369-388.

[12] 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.

[13] Dragsted, B., & Carl, M. (2013). Towards a classification of translation styles based on eye-tracking and keylogging data. Journal of Writing Research, 5(1).

[14] Van Waes, L., Leijten, M., & Quinlan, T. (2010). Reading during sentence composing and error correction: A multilevel analysis of the influences of task complexity. Reading and Writing, 23(7), 803-834.

[15] Raimes, A. (1987). Language proficiency, writing ability, and composing strategies: A study of ESL college student writers. Language Learning, 37(3), 439-468.

[16] Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. Journal of second language writing, 9(3), 259-291.

[17] Bipp, T., Lepper, A., & Schmedding, D. (2008). Pair programming in software development teams–An empirical study of its benefits. Information and Software Technology, 50(3), 231-240.

[18] Goncher, A., Johri, A., Kothaneth, S., & Lohani, V. (2009, October). Exploration and exploitation in engineering design: examining the effects of prior knowledge on creativity and ideation. In Frontiers in Education Conference, 2009. FIE'09. 39th IEEE (pp. 1-7). IEEE.

[19] Dorst, K., & Cross, N. (2001). Creativity in the design process: co-evolution of problem–solution. Design studies, 22(5), 425-437.

[20] Jacobs, J. K., Kawanaka, T., & Stigler, J. W. (1999). Integrating qualitative and quantitative approaches to the analysis of video data on classroom teaching. International Journal of Educational Research, 31(8), 717-724.

[21] Koschmann, T., Stahl, G., & Zemel, A. (2004, June). The video analyst's manifesto:(or the implications of Garfinkel's policies for the development of a program of video analytic research within the learning sciences). In Proceedings of the 6th international conference on Learning sciences (pp. 278-285). International Society of the Learning Sciences.

[22] Knoblauch, H., & Tuma, R. (2011). Videography: An interpretative approach to video-recorded micro-social interaction. The SAGE handbook of visual research methods, 414-430.

[23] Powell, A. B., Francisco, J. M., & Maher, C. A. (2003). An analytical model for studying the development of learners' mathematical ideas and reasoning using videotape data. The journal of mathematical behavior, 22(4), 405-435.

[24] Bezemer, J., & Mavers, D. (2011). Multimodal transcription as academic practice: A social semiotic perspective. International Journal of Social Research Methodology, 14(3), 191-206.

[25] Fonteyn, M. E., Kuipers, B., & Grobe, S. J. (1993). A description of think aloud method and protocol analysis. Qualitative Health Research, 3(4), 430-441.

[26] Langley, A. (1999). Strategies for theorizing from process data. Academy of Management review, 24(4), 691-710.

[27] Knigge, L., & Cope, M. (2006). Grounded visualization: integrating the analysis of qualitative and quantitative data through grounded theory and visualization. Environment and Planning A, 38(11), 2021-2037.

[28] Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R., ... & Sherin, B. L. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. The Journal of the Learning Sciences, 19(1), 3-53.

[29] Gu, Y. (2014). To code or not to code: Dilemmas in analysing think-aloud protocols in learning strategies research. System, 43, 74-81.

[30] Tan, C. T., Leong, T. W., & Shen, S. (2014, April). Combining think-aloud and physiological data to understand video game experiences. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 381-390). ACM.

[31] Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. Metacognition and learning, 9(2), 161-185.

[32] Snell, J. (2011). Interrogating video data: systematic quantitative analysis versus micro‐ethnographic analysis. International Journal of Social Research Methodology, 14(3), 253-258.

[33] Mannay, D. (2010). Making the familiar strange: can visual research methods render the familiar setting more perceptible?. Qualitative research, 10(1), 91-111.

[34] Park, K., & Kinginger, C. (2010). Writing/thinking in real time: Digital video and corpus query analysis. Language Learning & Technology, 14(3), 31-50.