A major impediment to investigations of writing has been the lack of a methodology that, on one hand, has the power to examine the larger patterns and fluctuations of complex cognitive processes and, on the other hand, provide the fine-grained detail necessary for understanding those processes. Although existing methodologies have provided many important insights into writing processes, those methodologies have not allowed researchers to identify the critical online cognitive processing that occurs during writing. Therefore, the purpose of this research was first to provide a brief review of existing writing methodologies; then to describe our new methodology that provides advantages over existing writing methods; and, finally, to demonstrate how we have used it to investigate the use of writing during problem solving.

Methodology used to study writing
In the early 1980s, Hayes and Flower introduced the use of the think-aloud protocol to investigate writing processes. To the present day, it is still used in writing research. Although originally there were skeptics about the use of verbal data as a window on cognitive processing (e.g. Nisbett and Wilson, 1977), Ericsson and Simon (1980) showed that concurrent verbal reports can be both accurate and relatively benign in their influence on the cognitive processes being reported. In traditional verbal-protocol analysis, large amounts of verbal data are recorded, transcribed, and then segmented into units of analysis that conform to a theoretical model. Hayes and Flower’s (1980) theoretical model of writing was developed using the content of think-aloud protocols. These think-aloud protocols revealed distinct writing processes, for instance, planning, translating, and reviewing, which occurred recursively during a writing task. For at least a decade, the think-aloud protocol was the method of choice in writing research, and the Hayes and Flower model was the theory of choice for many researchers.

TRAKTEXT: Investigating writing processes using eye-tracking technology

Douglas J Hacker, Matt C Keener and John C Kircher

Abstract
The purpose of this research was to provide a brief review of existing writing research methodologies, describe a new methodology in the investigation of writing, and demonstrate how this new methodology can be used in a pilot study to investigate the use of writing during problem solving. Our new methodology, TRAKTEXT, makes use of eye-tracking technology, which provides continuous measures of processing time, attention, and effort; does not disrupt the writer from the main task; produces data reflecting attentional shifts in periods of time as short as 100 milliseconds; can pinpoint text production or revision at the word level; and provides a more natural way of examining writing behaviors. In our exploratory study, we identified six unique writing behaviors. Results from the pilot study showed that writers who experienced a change in knowledge during problem solving demonstrated different writing behaviors from writers who did not experience a change in knowledge. Although TRAKTEXT provides several advantages over existing writing research methodologies, there are some components of writing (e.g. planning) that must be inferred from processing time and cognitive effort measures. Future iterations of TRAKTEXT may resolve these issues.

Keywords
Writing, eye tracking, cognitive effort, writing methodology

Department of Educational Psychology, The University of Utah, Salt Lake City, UT, USA

Corresponding author:
Douglas J Hacker, Department of Educational Psychology, The University of Utah, 1721 Campus Center Drive, Room 3220, Salt Lake City, UT 84112, USA.
Email: doug.hacker@utah.edu
Methodological Innovations

However, writing is a temporally and cognitively dense process, and verbal protocols do not provide a good understanding of the time course and cognitive effort of the processes identified by Hayes and Flower (Levy and Ransdell, 1996). Knowing how writing processes interact and change across time and how cognitive effort fluctuates with these processes can do much to support and develop theory. Hayes and Flower (1980) acknowledged early on that verbal protocols are incomplete and that “Many of the psychological processes on which we depend every day are completely unconscious and therefore will not appear explicitly in thinking-aloud protocols” (p. 216). Some processes are conscious at one point in the writing process and unconscious at other times, and this is particularly true for expert writers, who likely have automatized many of the subprocesses of writing. In addition, processes that occur rapidly are likely to be overshadowed in a verbal protocol and perhaps overlooked in favor of processes that require greater verbal description.

Finally, contrary to findings reported by Ericsson and Simon (1980), Janssen et al. (1996) showed that in a writing think-aloud condition, some pause times were significantly lengthened in relation to others, suggesting that the verbalization of some mental processes was intrusive to the processing occurring during writing.

Ronald Kellogg (1986, 1987a, 1987b, 1988) introduced a triple-task method for investigating writing processes that consisted of (a) the primary task of writing a text, (b) a secondary response time (RT) task that required the detection of an auditory signal, and (c) a directed retrospection task. For the directed retrospection task (Ericsson and Simon, 1980), participants are trained to identify their thoughts in terms of the three categories of writing that were identified by Hayes and Flower, planning, translating, and reviewing, in addition to an “other” category. With the secondary-task RT technique (Kahneman, 1973), participants are interrupted on a variable schedule, ranging from 15 to 45 seconds, by an auditory signal while writing, and they must say “stop” as soon as the auditory signal is detected. The RTs are recorded, with longer RTs indicating greater cognitive effort. Subsequent to saying “stop,” the participant presses one of the four buttons to indicate which of the four categories they were involved in at the time of the auditory interruption.

Kellogg’s triple-task method has several advantages over the think-aloud protocol. The directed retrospection indicates (a) which writing process is being used, (b) the time course of the three writing processes, and (c) the proportion of time devoted to each writing process. The RTs indicate the degree of cognitive effort being expended with each writing process. Moreover, the method allows for the investigation of many participants in an experiment in contrast to verbal-protocol analysis, which is a time-consuming process that works well for detailed analyses of a few participants but precludes the investigation of many participants.

However, there are shortcomings to Kellogg’s methodology. The auditory interruptions that are imposed on writers every 30 seconds or so can be very intrusive and disruptive to the writing process. In addition, after an interruption, the time participants need to resume writing is not reported. At the very least, several seconds must pass before the resumption of a writing process, and one must question whether the nature of the writing process has been changed because of the interruption. Having to choose among four a priori theoretical categories of writing also could produce a biased picture of writing (Janssen et al., 1996; Levy and Ransdell, 1996). Planning, translating, and reviewing can occur in parallel, and choosing among them can result not only in inaccurate identification of processes but in an increase in the selection of the fourth “other” category. This indeterminacy may be reflected in Kellogg’s (1987a) study in which he reported that 73% of the participants’ categorizations of their writing processes matched the experimenter’s categorizations. For writing research, this level of agreement is actually quite good; however, it does leave 27% of the writing processes in the “other” category. In another study, Kellogg (1987b) reported that 50% of writing was identified as translating, and 30% of writing was identified as planning and translating. This leaves 20% of the writing time designated as “other.”

Work by Piolat et al. (1999) has shown that when participants received more extensive training in identifying the three writing categories, they responded faster in selecting a category, which suggests that cognitive effort may fluctuate simply on the basis of the amount of training and not the writing process that is being engaged. Piolat et al. also found that the choice of the interruption interval might affect writers’ performance as a function of writing expertise: Younger or less experienced writers could exhibit cognitive overload with shorter intervals.

Kellogg (1987a) recognized that

During any given minute of composition, many or few words may be produced; composition probably proceeds by spurts and pauses more often than by a smooth rate of production. An analysis of how production rates fluctuate and of the function of pauses is an interesting issue. (p. 273)

Unfortunately, the triple-task method cannot produce this fine-grained detail on writing events that occur within seconds, nor can it provide records of the “spurts and pauses” that likely characterize writing. And although revision can be readily identified by a writer and confirmed by the changes that are made to the text, reviewing cannot be so readily disentangled from other writing processes (Levy and Ransdell, 1996). Reviewing of text could be an opportunity for planning, revision, or verifying that what is written is what was intended. Kellogg’s method did advance the scientific study of writing, but it also identified critical aspects of writing that could only be investigated using more advanced methodologies.

Levy and Ransdell (1996) introduced a process-oriented protocol analysis method that attempted to identify writing...
processes as they occur in real time. Their method is something of a combination of the think-aloud protocol of Hayes and Flower and the triple-task method of Kellogg; however, it introduces the use of a simple word processing program on a computer and monitor that participants use to write. In addition, Levy and Ransdell make use of two other computer programs, one that videotapes the participant’s writing as it appears on a computer monitor and a keystroke-capturing program. The video recording allows researchers to record the onset and offset of writing activities, and the keystroke-capturing program records every keystroke and the time at which it occurs, which allows for the recording of typing speeds, duration, and location of pauses. As participants write, they concurrently produce a think aloud that is audio recorded, and they respond to an auditory signal that occurs at a 30-second variable interval using a footswitch to record RTs.

Utilizing data from these multiple sources, Levy and Ransdell can provide converging evidence on a variety of writing processes. The word processing program identifies the text that is produced and revised. The audio recording of the think-aloud protocol provides insight into the thought processes underlying the text production or revision. The video and keystroke recording allow for the identification of discrete writing processes and the duration and frequency of those processes. And the RT data align measures of cognitive effort with specific writing processes that are empirically derived rather than theoretically imposed.

Levy and Ransdell (1994) argue that “Temporally dense process information is sorely needed in written composition research, and there has been only limited progress” (p. 222). Their process-oriented protocol analysis method does much to provide this information. However, their methodology suffers in many ways from similar shortcomings that were identified for the Hayes and Flower think-aloud protocol and for Kellogg’s triple-task method. In the Levy-Ransdell method, a writer must engage in a variety of writing processes (e.g., composing, reviewing, and planning), produce a concurrent verbal think aloud, and respond to a tone to record RTs. Arguably, the juxtaposition of these cognitive and behavioral processes could lead to high cognitive demands that potentially alter the writing process. As with all verbal protocols, the verbal-protocol data collected in this method suffer from incompleteness. Processes that occur rapidly are likely to be overshadowed and perhaps overlooked by processes requiring greater verbal description. Whether rapid processing would be detected by Levy and Ransdell’s multi-pronged method has not been established. Finally, the auditory interruptions occurring every 30 seconds or so to record secondary RTs are intrusive and disruptive. This, in addition to the time needed to resume writing, could alter the natural flow of the writing process.

Piolat et al. (1999) introduced a computer program called SCRIPTKELL, which is a variation of Kellogg’s triple-task procedure. The researchers state that the program is not necessarily restricted to writing research and can be used to measure cognitive effort and processing time in other areas, such as reading, learning, and problem solving. Focusing on its applications to writing research, the SCRIPTKELL program allows the researcher to manipulate many of the parameters that are associated with the triple-task procedure. For example, an experiment can be configured in a way that allows the researcher to specify the writing task, writing time, auditory signals and timing of the auditory signals, and secondary RT task. Also, rather than being restricted to four writing categories from which the participant must select in the directed retrospection, the program allows for the selection of up to 12 categories that can be specified by the researcher.

In sum, SCRIPTKELL is an automated version of the Kellogg’s triple-task procedure that provides greater flexibility in the kinds of processes that can be investigated in terms of cognitive effort and processing time. Regrettably, in the context of writing, it suffers from the same shortcomings of Kellogg’s original methodology. The auditory interruptions that are imposed on writers to measure RTs can be intrusive and disruptive to the writing process, and the time interval may not be sufficiently short to record the temporally dense process information characteristic of writing. In addition, after an interruption, the time participants need to resume writing is not reported. And, although the number of writing processes from which the participants can choose is expanded, they still remain a prior theoretical categories of writing. Thus, the writing processes identified are theoretically imposed rather than empirically driven.

Beers (2004) introduced a methodology that makes use of eye-tracking technology. For reasons we will describe below, we believe that the use of eye-movement measures is an important approach to the study of writing; however, the focus of Beers’ work was primarily on reading during writing and not writing per se. Beers produced interesting results concerning how participants used reading during the writing process, but he did not have much to say about writing.

Alamargot et al. (2006) also developed a methodology that makes use of eye-tracking technology in the investigation of reading and writing processes. Their computer program called Eye and Pen provides many advancements over Beers’ methodology in the analysis of writing. The Eye and Pen software yokes signals from a digitizing tablet with the signals from an eye tracker. Yoking these two signals together makes possible investigations of textual information fixated during reading of a text and a person’s handwriting as text is produced on the tablet. Although the program is dedicated to the study of handwriting, it does allow for the online investigations of the dynamics between a person’s reading of textual information, including the textual information that has been written by the participant, with the participant’s writing. The data collected by the Eye and Pen program allow for the analysis and reconstruction of each word the participant writes and all pauses that he or she makes during a reading/writing session. Therefore, similar to Levy and Ransdell’s (1996) writing
methodology in which the participant’s writing is videotaped as it appears on a computer monitor, the Eye and Pen program can replay every gaze, saccade, or fixation as a participant produces or reads written text.

In two studies, the utility of Eye and Pen has been demonstrated. Alamargot et al. (2007) used the Eye and Pen program to demonstrate that parallel processes during reading/writing differ depending on the proximity to a writing pause and that the frequency and duration of parallel processing depends on the writer’s cognitive capacity. In addition, Alamargot et al. (2012) reported using the Eye and Pen program to understand the nature of processing involved in the production of subject–verb agreement.

Alamargot et al. have produced a versatile methodology for the study of reading and writing that provides for a fine-grained analysis of the dynamic and continuous online processes that are brought to bear in the interplay of complex cognitive processes during reading and writing. Their Eye and Pen program makes parallel the investigation of writing processes that can occur in periods of time as short as milliseconds or as long as several seconds and can analyze discrete data, such as fixations and pauses, or continuous data that can provide insights into the acceleration or deceleration of writing and eye movements. Alamargot et al. (2006) do acknowledge that future studies are needed to investigate the nature of pause times. More specifically, writers can remain fixated at a specific portion of text and yet their attention may be directed to other things, such as what else to write or how to rephrase something that has already been written, or to thoughts that are completely decoupled from the text as is the case with mindless reading (Smallwood and Schooler, 2006). Alamargot et al. state that “it is difficult to state with certainty that every item of fixated visual information is an item of processed information” (p. 296). We are in full agreement with this statement and offer as a possible remedy the use of pupil diameter as a measure of cognitive engagement during these pause times. The degree to which a writer is cognitively engaged during pause times could help to infer the kind of processing that is occurring.

**TRAKTEXT**

Although existing writing research methodologies have increased our understanding of writing, with the exception of the Eye and Pen device, they do not reveal the moment-to-moment, time-course processes involved during writing. Shortcomings of existing methodologies include problems with retrospective reports; the course granularity of secondary RTs; occasional lack of specificity regarding which writing process is occurring; unnatural writing experiences; invasiveness and disjointedness in providing a cohesive representation of writing; rapidly occurring writing processes can be overshadowed by more dominant processes; large percentages of component writing processes are left unidentified; writers are asked to identify their writing processes by selecting from a group of a priori categories of writing behaviors that are theoretically imposed rather than empirically derived; writers also may be forced to choose among several processes occurring simultaneously; and finally, lack of a method of measuring online the cognitive effort that a writer expends during text production or revision.

As an alternative to existing research methodologies, we have developed a new methodology to investigate writing that we believe addresses many of these shortcomings. Our new methodology makes use of eye-tracking technology, which, although not problem free, is considered the state of the art for online investigations of reading (Boland, 2004; Carreiras and Clifton, 2004; McConkie et al., 1979; Rayner, 1998). Eye tracking provides continuous measures of processing time, attention, and effort; does not disrupt the participant from the main task; produces data reflecting attentional shifts in periods of time as short as 50 milliseconds; can pinpoint comprehension problems at a word and even intra-word level; and provides a more natural way of examining word and sentence processing (Boland, 2004; Carreiras and Clifton, 2004; Mitchell, 2004; Pickering et al., 2004).

Although there are many obvious similarities concerning the use of eye tracking for reading and writing, there are some challenging differences. Investigations of reading make use of static stimuli that can be easily deconstructed into sentence, clausal, word, or sub-word units of analysis. In contrast, writing creates dynamic stimuli that change with every change the writer makes, and these stimuli changes must be simultaneously synchronized with the moment-to-moment movement of the writer’s eyes. We designed TRAKTEXT to track the dynamics of writing and capture online the moment-to-moment, time course of eye movements associated with them.

TRAKTEXT can display several unique screens on a computer monitor. Screens can display static stimuli such as textual information, graphics, or other visual stimuli. Using traditional eye-tracking techniques, participants’ reading behaviors can be tracked while they read these static stimuli. In addition, screens can be designed for participants to write their responses using the keyboard. Participants can freely navigate among the reading and writing screens, and regardless of which screen they view, eye measurements are continuously recorded. With the writing screen, the dynamic nature of writing is captured by generating and recording a bitmap of the screen contents each time the participant strikes the space bar or return key or moves the cursor to other parts of the text. Each time a bitmap is created, a marker is sent to the eye-tracking program so that we can synchronize the spatial coordinates for each currently active word on the writing screen with the coordinates recorded by the eye-tracking technology. In this way, we can associate each word that has been added, edited, or deleted from the text with specific eye measurements. A full illustration of how TRAKTEXT records writing will be described in the method section of our pilot study.

Methodological Innovations
TRAKTEXT records numerous measures that provide fine-grained detail of the behaviors associated with the writing process and product. Each word written, deleted, or edited is recorded along with the amount of time spent writing, reading, or rereading each. Cognitive effort allocated to the writing and reading of each word is measured using pupil diameter as an indicator. Numerous studies have shown that cognitive demand is associated with pupil dilation (e.g., Ahern and Beatty, 1979; Granholm et al., 1996; Just and Carpenter, 1993; Kahneman and Beatty, 1966). Kahneman (1973) has shown that pupillometric response is a reliable and sensitive psychophysiological index of the momentary processing load during performance of a wide variety of cognitive activities, including attention, memory, language, and complex reasoning. Moreover, pupillometric response shows within-task, between-task, and between-individual variations in processing load and cognitive resource capacity (Beatty, 1982; Kahneman, 1973).

A pilot study using TRAKTEXT

Now that we have presented an overview of writing methodology and described TRAKTEXT, we turn our attention to using TRAKTEXT in a pilot study to demonstrate how the program can be used to investigate writing research questions. In our pilot study, we first identified unique writing behaviors by observing participants online as they wrote. For each identified behavior, we measured processing time and pupillary changes. Once we identified writing behaviors, we examined how participants used these writing behaviors in a problem-solving task. The problem we presented to them involved the resolution of a misconception concerning gravity. Our research question was, ‘Do the writing behaviors used by participants who undergo a knowledge change differ from the writing behaviors used by students who do not experience a knowledge change?’ Our hypothesis was that participants who undergo a knowledge change will use writing in ways that reflect knowledge transformation in contrast to participants who do not undergo a knowledge change and use writing in a simple knowledge telling fashion (Bereiter and Scardamalia, 1987).

Method

Participants

Participants were 15 freshman, sophomores, or juniors enrolled in a college introductory educational psychology course. The students enrolled in the course either were currently admitted to a teacher education program or were about to apply to one. Students were predominantly white and middle class. Participation in the study satisfied the research participation requirement for the course. Alternative ways to satisfy the requirement were provided.

Apparatus

Participants’ eye movements were monitored using an Applied Sciences Laboratory (ASL) model 501 head-mounted eye tracker. The ASL head mount tracks head movements and adjusts pupil and fixation measures with each movement. The eye tracker was interfaced with two 1.8-GHz Hewlett Packard desktop computers: One ran the eye tracker and recorded the data, and the other ran the experiment. Participants had freedom of head movement while wearing the eye tracker. Viewing was binocular, and eye movement was recorded from each participant’s right eye. The eye tracker recorded fixation samples at 60 Hz, approximately one sample every 17 milliseconds. Participants were seated directly in front of the computer monitor, and the approximate distance between the right eye and monitor was 25 in.

Oculomotor measures

Eye fixations. A total of four criteria were used to define a fixation (Eye Tracker System Manual, ASL EyeTrac 6, 2007). First, a fixation began at the first of six consecutive samples that occurred within 0.5° of visual angle. Second, any three consecutive fixation samples farther than 1° of visual angle in the horizontal or vertical direction from the running mean position ended the fixation. Third, the final fixation position was the mean position of all fixation samples between the beginning and end of the fixation period, but any two or fewer consecutive fixation samples that were farther than 1.5 standard deviations from the mean position were excluded from the calculation of the final position. Fixation samples were recorded at 17 milliseconds; therefore, the shortest duration of an eye fixation was approximately 100 milliseconds or 0.1 second. Following eye-tracking conventions established in the reading literature, any eye fixation duration longer than 1 second was considered an artifact and deleted. Of the 15,686 fixations recorded in this study, 278 or 1.8% were longer than 1 second and deleted from the analyses. For any one participant, the longest average length of fixations over 1 second was 1.66 seconds.

Pupil diameter. For each fixation sample, a pupil diameter was measured using the number of pixels on the eye camera that spanned the diameter of the participant’s right pupil. This number will change with different eye cameras, but all participants used the same apparatus for this study. The pupil diameter for an eye fixation was calculated as the mean pupil diameter of all fixation samples constituting that eye fixation. Although measures of pupil diameter can be affected by gaze direction, given the short distance from the right pupil to the middle of the monitor (i.e. approximately 25 inches) and the centered position of the participant in front of the monitor, we believe that fluctuations due to gaze direction were minimized.
Materials

Pre-problem task. TRAKTEXT starts with a blank screen with a simple toolbar at the top that allows the participants to use a mouse to click on a “Text,” “Task,” or “Write” button that navigates them to the corresponding screens. There are other buttons on the menu bar for the experimenter to stop and start the program; create new files, folders, or projects; and save the data. These buttons are not accessible to the participants once they begin an experiment.

The pre-problem involved a simple reading and writing task that concerned the Jack and Jill nursery rhyme. Using a mouse, participants began by clicking on the “Text” button, which presented this statement in the middle of the screen, “Remember the famous nursery rhyme about Jack and Jill?” Participants then clicked on the “Task” button, which presented the statement in the middle of the screen, “Write your description of what happened to Jack and Jill as completely as you can.” They then clicked on the “Write” button and began to write their answers to the task. All writing utilized Courier, a fixed-width font. Participants could take as much time as they wished; revise their answers as much as they wished; and could navigate back and forth to the “Text,” “Task,” or “Write” screens. Once finished with the pre-problem, participants were asked whether they were familiar with the nursery rhyme, and all indicated that they were.

There were two purposes to the pre-problem task. First, reading and writing on the pre-problem helped participants adapt to the use of the TRAKTEXT program and adjust to the eye-tracking equipment. Second, because of the simplicity of the task, we used participants’ reading and writing to calculate baseline measures for both processing time and pupil diameter. We calculated four baseline measures: writing rate (millisecond/character), writing pupil diameter, reading rate (millisecond/character), and reading pupil diameter.

Writing task. In contrast to the writing tasks that are typically presented to participants in writing studies (e.g., write a persuasive essay, develop and write arguments for and against some issue, or write an article for a school newspaper), we asked participants to use writing while problem solving. The problem we presented to participants involved a common misconception about the effects of gravity on falling objects. When participants clicked on the “Task” button, the problem was presented:

Two jugglers in a circus were each balancing a bowling ball on his nose. One juggler was speeding around the circus ring while standing in the back of a wagon, and the other was standing still in the center of the ring. Whose bowling ball would hit the ground or floor of the wagon first, provided that the distance to fall was the same for the two jugglers, and they both dropped the balls at the same time? Why? Write your answer and your explanation in complete sentences as if you were to submit them for a grade in a science class.

This problem has been shown to reveal people’s misconceptions about the dynamics of gravity, and consequently, many people initially get the answer incorrect (Guzzetti et al., 1993). This was later confirmed in our analyses in which we found that 12 of our 15 participants initially wrote an incorrect answer.

Participants were given as much time as they wished to read the problem. Once finished, they then clicked on the “Write” button and typed their answers. Participants could freely write, delete, and edit their answers until they were satisfied with their responses. At this point, participants clicked on the “Text” button, and the following text appeared. The text contained information necessary to answer the question correctly:

Although people have a hard time believing it, the horizontal component of motion for a projectile is not tied to its vertical component of motion. Each acts independently of the other. Because they act independently, the path of a projectile is a curve. The path is a curve because the projectile keeps on moving forward at a constant rate, just as it would be moving if it were not dropping, and it keeps on dropping at an accelerating rate, just as it would if it were not moving forward. The projectile’s forward motion is not affected by its downward motion, and its downward motion is not affected by its forward motion.

If participants wrote an incorrect answer, this text served as a refutational text, that is, the text provided information that confronted them with their misconceptions about the topic of gravity and presented alternative conceptions that could be used to repair their misconceptions (Hynd and Guzzetti, 1998). The use of refutational text has been shown to be an effective tool in resolving students’ misconceptions in areas of science (Guzzetti et al., 1993). If participants knew the correct answer and wrote it at the outset, this text served only as an informational text.

Procedure

After filling out a consent form, each participant was seated in front of the computer monitor, fitted with the eye tracker, and calibrated to the tracker. We first presented participants with the Jack and Jill pre-problem exercise using the TRAKTEXT program. When finished with the pre-problem, participants were presented with the gravity problem. First, they were presented with the task, which they could read as many times as they wished before starting to write. They then navigated to the writing screen to write their response to the problem. During writing, they could navigate back to the problem whenever they wanted, and they were free to revise their answer in any way. When finished with their answers, participants selected the informational/refutational text, which could be read as many times as they wished. Once finished reading, they navigated back to the writing screen to revise their original answer in any way they
deemed necessary, and they could take as much time as they wanted. Also, during the revision of their answers, they were free to navigate to the task or informational/refutational text. Regardless of whether participants produced a correct or incorrect answer, they were given no hints or prompting by the researcher. When they finished writing, they were asked whether they were satisfied with their answer, and if so, they were dismissed from the laboratory.

Results

Data recording

The ASL eye tracker produces an EHD file that we converted to an ASCII file. TRAKTEXT uses this ASCII file to generate a fixation file that records a bitmap each time the space bar or return key is struck or the cursor is navigated to different parts of the text. The fixation file contains a record of each word written, deleted, or edited in the written text and provides a chronological ordering of the oculomotor data for each fixation that occurred during reading of the problem task, the informational text, and for words written for each bitmap. EYEDIT, another program we created, uses the fixation file to play back chronologically the entire reading and writing processes and displays eye fixations that occurred during reading and writing. The number of fixations displayed can be varied from one up to the total number of fixations that occurred during a participants’ session. As the fixations are displayed, their positions can be edited.

Baseline measures

Writing baseline measures. We used the pre-problem task to calculate four baseline measures: writing pupil diameter, writing rate (millisecond/character), reading pupil diameter, and reading rate (millisecond/character). For the writing baseline measures, we used pupil diameter as a measure of cognitive effort required for word production plus the graphomotor activity for keyboarding them. Using the fixation files that recorded the Jack and Jill exercise, we identified bitmaps that recorded words written but omitted the first five words and last three words. These words were omitted because the first few words written may have required additional effort and time to produce as participants initiated and oriented to the task, and the last few words may have required less effort and time to produce as participants completed the task. Using the fixations that fell on these middle words, we selected the four lowest pupil diameter measures and averaged the four to arrive at a baseline pupil diameter for reading. We also calculated the total amount of time a participant spent on these fixated words, counted the number of characters, and derived a reading base rate expressed as millisecond per character.

Identification of writing behaviors

For the problem task and text, we recorded traditional reading measures (e.g. words fixated, number of fixations, and time per fixation) as the participants read these two texts. Although we realize that an in-depth analysis of participants’ reading of the texts could provide valuable insights into how reading content contributes to the writing content, for the purposes of this study, we did not focus on those texts but instead focused exclusively on writing and reading behaviors that we observed on the bitmaps that were generated while the participants wrote their answers to the gravity problem.

Using the fixation file that was generated for each participant during the problem-solving task, we identified for each bitmap the fixations that were recorded when the bitmap was generated. The EYEDIT program then permitted us to synchronize the word recorded on the bitmap with the eye fixations recorded during that time frame. From these fixations, we calculated mean pupil diameter and total time for each word. Each bitmap was analyzed in this way. Therefore, rather than using theoretically driven conceptualizations or self-reports of writing processes, we empirically identified writing processes directly from the observed writing behaviors. We identified seven behaviors. Table 1 illustrates the time course of these writing behaviors along with the pupil diameter and cognitive effort allocated to each behavior for one of our participants.

Area of monitoring and production. We defined an area of monitoring and production (AMP) as all the words that a participant wrote before he or she stopped writing and made a fixation that extended beyond 12 characters or spaces from the point of word production. Because there are a number of implicit and explicit monitoring processes involved during
Table 1. Time course of writing behaviors and cognitive effort allocated to each behavior during writing answer to the gravity problem.

| Participant | TSK-1  | AMP-1  | AMP-2  | ARF-1  | AMR-1  | AMP-3  | ARF-2  | AMR-2  | AMP-4  | ARS-1  | TXT-1  | AMR-3  | AMP-5  | ARS-2  | AMP-6  | ARS-2  |
|-------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Total time  | 91,683 | 42,744 | 3066   | 22,629 | 1383   | 3233   | 733    | 33,030 | 2750   | 58,316 | 38,000 | 21,196 | 2650   | 22,329 | 11,100 |
| Fixated on words | 27,450 | 2950   | 14,383 | 7167   | 533    | 1966   | 0      | 22,549 | 1700   | 18,200 | 2583   | 16,399 | 6184   |
| Proportion of time fixed on words | 0.64   | 0.96   | 0.64   | 0.47   | 0.38   | 0.61   | 0      | 0.68   | 0.62   | 0.84   | 0.86   | 0.98   | 0.73   | 0.56   |
| PD (SD)     | 53.18 (3.87) | 54.16 (1.26) | 49.56 (2.95) | 45.61 (6.58) | 50.72 (1.90) | 51.62 (1.56) | 0      | 49.20 (4.02) | 44.95 (2.70) | 48.55 (3.11) | 47.56 (2.56) | 53.37 (19.3) | 48.62 (15.18) | 45.02 (3.03) |
| Cognitive effort | 1.08   | 1.21   | 1.01   | 1.02   | 1.03   | 1.05   | 0      | 1.00   | 1.00   | 0.99   | 0.97   | 1.19   | 0.99   | 1.00   |
| Extra time  | 17,352 | 1553   | 10,485 | −9089  | 783    | 1853   | −217   | 15,366 | 176    | −2200  | 1600   | 1059   | 6597   | −96.91 |
| Predicted time on behavior | 25,392 | 1513   | 12,144 | 24,289 | 600    | 1380   | 1000   | 17,664 | 2574   | 40,200 | 19,596 | 1591   | 15,732 | 20,791 |
| Correct     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 1      | 1      | 1      | 1      | 1      | 1      |
| Words written (AMP), deleted (AMR), reviewed (ARF/ARS), or both (ARMR/AMPR) | the man who was in the back of the truck; probably have his howling (sic) ball hit the floor; truck in moreof have moreof (sic) because have because the would moreof moreof have moreof probably moreof the because moreof probably the; pushing it downward—the speed of the truck as well as gravity; who back was; l gravity as well as truck the of speed the downward—it pushing more have would it because first; truck the of floor the hit ball howling his have probably would truck the of back the in was who man; balls would likely hit the ground at the same time because they're the same weight and it's the same distance to the ground.

Final answer: The two balls would likely hit the ground at the same time because they're the same weight and it's the same distance to the ground.

TSK: task; AMP: area of monitoring and production; AMR: area of monitoring and revision; ARF: area of review—formative; ARS: area of review—summative; TXT: text; ARMR: area of review, monitoring, and revision; AMPR: area of monitoring and production with review; SD: standard deviation.
normal language production (Postma, 2000), we labeled these areas of text as involving both monitoring and production. AMPs are recorded as one or more contiguous bitmaps, each one containing a written word along with the location and timing of every fixation that occurred during the production of the bitmap. Pupil diameter is averaged across all fixations that occurred on a fixated word and across all the bitmaps that occurred in an AMP. Total time a participant fixated on a word, total time for the duration of a bitmap, and total time for an AMP are recorded.

For the participant illustrated in Table 1, he or she took 91,683 milliseconds to read the task and then proceeded to write, producing an AMP of 25 bitmaps that contained the words, “the man who was in the back of the truck would probably have his howling (sic) ball hit the floor.” The participant then made a fixation of >12 characters or spaces from the last word produced to review an earlier part of his or her written text, “the man who” (see section “Area of review—formative”). The average pupil diameter was 53.18, and the total time to produce this AMP was 42,744 milliseconds. Average pupil diameter for the AMP is divided by the writing baseline pupil diameter produced in the Jack and Jill task to produce a ratio of cognitive effort of 1.08, which indicates the participant expended more cognitive effort writing this AMP than what he or she expended when writing the Jack and Jill task. Dividing the total time by the number of characters and spaces written in the AMP provides a keyboarding time that is compared to the baseline keyboarding time from the Jack and Jill task. In this case, based on the participant’s writing for the Jack and Jill task, the amount of time the participant should have taken to produce the words written for the bowling ball problem was 25,392 milliseconds; however, the participant used 42,744 milliseconds, which indicates that the participant used an additional 17,352 milliseconds during this AMP. This additional amount of time will be discussed further in the following section.

Two additional measures provide other insights into how the words were written. Of the 42,744 milliseconds the writer took to produce this AMP, only 27,450 milliseconds of them were spent directly fixated on words for a ratio of 0.64. For AMPs, the amount of time not fixating on words is usually due to looking at the keyboard as the words are being typed. This ratio will vary by type of writing behavior, and more of this will be discussed later. However, pupil diameter can still be used as a measure of cognitive effort even though the words currently being typed are not being directly fixated. Pupil dilation under cognitive effort occurs even when the eye is not directly perceiving an object (Beatty, 1982).

There are three additional AMPs that occurred before the writer navigated to read the text. The second AMP entailed the production of “of truck first because it would have moreof (sic),” the third AMP “of a,” and the fourth AMP “pushing it downward—the speed of the truck as well as gravity.” After reading the text, the writer deleted the entire response and engaged in two more AMPs that indicated a change in his or her thinking about the problem: “balls would likely hit the ground at the same time because,” “the same weight and it’s the same distance to the ground.”

**Area of review—formative.** An area of review—formative (ARF) is defined as an area of text that is beyond 12 characters or spaces from the last point of word production and is being reread by the writer before he or she resumes writing. An ARF is recorded on a single bitmap on which fixations are found on words beyond the 12 characters or spaces criterion. Pupil diameter is averaged across all fixations that occur on a word and for all words during the review, and the total amount of time for the fixations on all words is recorded.

An ARF occurred in our example (see Table 1) when the writer returned to “the man who.” The writer spent 3066 milliseconds on this review, and cognitive effort as compared to the Jack and Jill task increased to 1.21. According to the base rate established for reading in the Jack and Jill task, the rereading of “the man who” should have taken 1513 milliseconds, which allowed for an additional 1553 millisecond on this review. For the first two recorded writing behaviors, the participant wrote a partial answer, made a quick three second regression to the beginning of the written text during which his or her cognitive effort increased. A second ARF occurred shortly thereafter that was considerably longer in duration and amount of text reread. During this second ARF, the writer engaged in reviewing and rereviewing portions of the answer. A third brief ARF of about 2.65 seconds occurred after he or she had changed the written response.

**Area of review—summative.** An area of review—summative (ARS) is similar to an ARF in that it is defined as an area of text that is fixated beyond 12 characters or spaces from the last point of word production; however, an ARS is a review process that occurs at the end of writing or revision, and no further changes are made to writing. At this point, the writer has produced text, makes a saccade to reread previously written text and then ends further text production or revision. An ARS is recorded on a single bitmap on which fixations are found on words beyond the 12 characters or spaces criterion. Pupil diameter and rereading time are recorded and calculated in the same ways as with ARFs.

In our example (see Table 1), the writer went through a series of writing behaviors, including multiple AMPs and ARFs, and did a final review of the written text, “who was back,” and then navigated to the text. The ARS was very brief, lasting 2750 milliseconds, which was only 176 milliseconds more than the reading base rate, and cognitive effort was at the base rate of 1.00, which was comparable to the cognitive effort for reviewing written text during the Jack and Jill task. A second ARS occurred at the very end of the participant’s session.

**Area of monitoring and revision.** An area of monitoring and revision (AMR) is defined as any change the writer makes to
his or her text (cf. Fitzgerald, 1987). Similar to word production, word revision can involve both implicit and explicit monitoring processes (Postma, 2000); therefore, as with AMPs, AMRs involve a monitoring process. AMRs are recorded as one or more contiguous bitmaps, each one indicating that text has been deleted. The deleted text can consist of single characters or spaces or entire words. Pupil diameter is averaged across all fixations that occur on the characters, spaces, or words that are deleted, and the amount of time for the fixations is recorded. The average pupil diameter for the AMR is divided by the writing baseline pupil diameter to produce a ratio of cognitive effort. The total time for the deletion is compared to a standard rate of deletion of 200 milliseconds per character or space, which was the average deletion rate for the computer used in our research. This rate can vary by computer and within the same computer depending on concurrent demands on the computer. The rate also can vary if the writer decides to hold down the deletion key for large sections of text or to hit the deletion key separately for each character or space. The 200 milliseconds per character or space that we used was the average across all these variations.

Returning to our example, the writer engaged only briefly in revision for the first answer. After an ARF of 15,200 milliseconds, the writer deleted “of,” then wrote “of a,” deleted “of a,” and then engaged in a more lengthy AMP. The two AMRs combined took about 2 seconds and low cognitive effort. For the second AMR, there were too few recorded fixations to calculate a reliable pupil diameter measure. Once the participant read the informational/refutational text, he or she engaged in a much lengthier AMR in which the entire answer was deleted.

**Area of review, monitoring, and revision.** In some of our observations of writing behaviors, we found that writers occasionally combined processes. One such instance is the combination of revision and rereading. This occurs when the writer holds down the deletion key while rereading portions of the text that are beyond 12 characters or spaces from the point of deletion. Each area of review, monitoring, and revision (ARMR) is analyzed as two separate behaviors consisting of an ARF and an AMR. Pupil diameter and time are recorded and calculated in the same ways as these two writing behaviors and then combined. In this study, ARMRs did not occur often, and no ARMRs occurred in the writing behavior illustrated in Table 1.

**Area of monitoring and production with review.** A second combination of writing behaviors that we observed involved text production and rereading. These areas of monitoring and production with review (AMPRs) occur when the writer is keyboarding new text and at the same time reviews existing text beyond the 12 characters or spaces criterion. AMPRs are analyzed as two separate behaviors consisting of an AMP and an ARF. Pupil diameter and time are recorded and calculated in the same ways as these two writing behaviors and then combined. AMPRs were also rare in this study, and there is no instance of one in the example illustrated in Table 1.

**Potential planning.** The amount of time for each of our observed behaviors was recorded, and using our baseline measures of text production, deletion, and reading, we calculated a predicted time for each behavior. For instance, the writer took 42,744 milliseconds to write the first AMP in our example (“the man who was in the back of the truck would probably have his howling (sic) ball hit the floor”). The AMP contains 92 characters and spaces, and based on his or her writing base rate of 276 milliseconds per character established in the Jack and Jill task, we calculated a predicted production time for the 92 characters and spaces of 25,392 milliseconds. The difference between actual and predicted times indicates that the writer took 17,352 milliseconds more to write this text than to write a comparable amount of text in the simple Jack and Jill task. Similar calculations were made for the other writing behaviors except for AMRs. Because AMRs involve the deletion of text, we calculated an average deletion rate for the computer that we used for the study.

All of the positive differences between actual and predicted times for each writing behavior represent additional time the writer spent on each behavior, and negative differences represent times when the writer went quickly through the behavior in relation to the simple Jack and Jill task. Unfortunately, we cannot know what writers are doing with this extra time, and because they are brief, lasting from less than a second to 17 seconds, trying to engage the writer in a verbal protocol to describe his or her thinking would not be practical. We can speculate that the writer could be engaged in considering what has been written and what needs to follow or evaluating whether the written text fully and accurately expresses what had been planned or simply evaluating his or her spelling. Hayes and Nash (1996) describe planning as “preparation for action” (p. 29), and because these periods of time may be described as preparations for further action, we have labeled them as potential planning. We recognize that this time may not be used for planning at all and may involve daydreaming or simply taking a mental break from the writing task. However, the amount of extra time for each writing behavior is usually brief and is usually followed by an intentional action, such as produce more text, delete text, or review text. Moreover, although this was not always the case, extra time was often associated with increased cognitive effort. Therefore, there is some prima facie evidence that the extra time is at the very least being directed to some aspect of continued word production or review.

**Overview of writing behaviors**

The descriptive statistics for the writing behaviors for each participant are given in Table 2. There was great variability
Table 2. Writing behaviors for each participant: number; mean time with standard deviation (SD), minimum (min), and maximum (max); and mean cognitive effort (CE) with standard deviation, minimum (min), and maximum (max).

| Participant no. | Area | No. of areas | Time mean | Time SD | Time min | Time max | CE mean | CE SD | CE min | CE max |
|----------------|------|--------------|-----------|---------|----------|----------|---------|-------|--------|--------|
| 1 trans AMP | 4 | 17,487 | 15,993 | 1533 | 36,960 | 1.00 | 0.04 | 0.96 | 1.04 |
| ARF | 4 | 8391 | 7253 | 1150 | 15,233 | 1.14 | 0.01 | 1.12 | 1.15 |
| ARS | 2 | 6458 | 2769 | 4500 | 8416 | 1.08 | 0.03 | 1.07 | 1.10 |
| AMPR | 0 | | | | | | | | |
| ARMR | 0 | | | | | | | | |
| 2 trans AMP | 6 | 24,193 | 13,239 | 3233 | 42,744 | 1.01 | 0.04 | 0.97 | 1.08 |
| AMR | 3 | 13,372 | 21,330 | 733 | 38,000 | 1.01 | 0.03 | 0.99 | 1.03 |
| ARF | 3 | 6972 | 7128 | 2650 | 15,200 | 1.14 | 0.11 | 1.02 | 1.21 |
| ARS | 2 | 6925 | 5904 | 2750 | 11,100 | 1.00 | 0.00 | 1.00 | 1.00 |
| AMPR | 0 | | | | | | | | |
| ARMR | 0 | | | | | | | | |
| 3 teller AMP | 4 | 17,064 | 20,610 | 2666 | 56,496 | 1.24 | 0.29 | 1.03 | 1.45 |
| AMR | 6 | 1694 | 1517 | 233 | 4566 | 0.91 | 0.04 | 0.84 | 0.94 |
| ARF | 9 | 4159 | 5040 | 400 | 17,000 | 1.01 | 0.08 | 0.89 | 1.20 |
| ARS | 2 | 3783 | 424 | 3483 | 4083 | 1.08 | 0.03 | 1.06 | 1.10 |
| AMPR | 0 | | | | | | | | |
| ARMR | 3 | 1521 | 1451 | 316 | 3133 | 1.06 | 0.05 | 1.00 | 1.11 |
| 4 trans AMP | 4 | 17,526 | 16,443 | 3900 | 40,375 | 1.06 | 0.03 | 1.02 | 1.09 |
| AMR | 2 | 11,366 | 15,014 | 750 | 21,983 | 1.04 | 0.14 | 0.97 | 1.16 |
| ARF | 3 | 10,166 | 2452 | 7350 | 11,833 | 1.05 | 0.01 | 1.04 | 1.05 |
| ARS | 2 | 14,816 | 2262 | 13,216 | 16,416 | 1.07 | 0.06 | 1.03 | 1.11 |
| AMPR | 0 | | | | | | | | |
| ARMR | 0 | | | | | | | | |
| 5 trans AMP | 7 | 11,648 | 7707 | 1533 | 19,460 | 1.08 | 0.03 | 1.04 | 1.12 |
| AMR | 2 | 22,258 | 30,416 | 750 | 43,766 | 1.07 | 0.14 | 0.97 | 1.16 |
| ARF | 4 | 7536 | 9111 | 233 | 21,116 | 1.06 | 0.07 | 0.98 | 1.15 |
| ARS | 2 | 14,858 | 1873 | 13,333 | 16,183 | 0.97 | 0.04 | 0.94 | 1.00 |
| AMPR | 0 | | | | | | | | |
| ARMR | 0 | | | | | | | | |
| 6 teller AMP | 2 | 12,930 | 15,410 | 2033 | 23,827 | 1.06 | | | |
| AMR | 0 | | | | | | | | |
| ARF | 1 | 2366 | | | | | | | |
| ARS | 2 | 5350 | 7283 | 200 | 10,500 | 0.96 | 0.14 | 0.86 | 1.06 |
| AMPR | 0 | | | | | | | | |
| ARMR | 0 | | | | | | | | |
| 7 trans AMP | 18 | 7892 | 6970 | 466 | 22,496 | 1.08 | 0.05 | 1.00 | 1.19 |
| AMR | 15 | 3446 | 9077 | 116 | 36,116 | 1.05 | 0.05 | 0.97 | 1.16 |
| ARF | 4 | 4719 | 5065 | 516 | 19,282 | 1.05 | 0.06 | 0.94 | 1.14 |
| ARS | 3 | 13,694 | 3620 | 10,166 | 17,400 | 1.04 | 0.01 | 1.03 | 1.04 |
| AMPR | 3 | 344 | 149 | 250 | 516 | | | | |
| ARMR | 6 | 1005 | 721 | 233 | 2299 | | | | |
| 8 trans AMP | 7 | 8131 | 5408 | 283 | 16,031 | 1.03 | 0.05 | 0.99 | 1.11 |
| AMR | 6 | 6344 | 7669 | 416 | 19,116 | 0.98 | 0.07 | 0.89 | 1.08 |
| ARF | 2 | 15,599 | 1744 | 14,366 | 16,833 | 1.01 | 0.05 | 0.97 | 1.05 |
| ARS | 1 | 15,816 | | | | | | | |
| AMPR | 0 | | | | | | | | |
| ARMR | 0 | | | | | | | | |
| 9 trans AMP | 4 | 9136 | 6160 | 166 | 14,215 | 1.01 | 0.02 | 0.98 | 1.03 |
| AMR | 2 | 2591 | 294 | 2383 | 2799 | 0.96 | 0.07 | 0.91 | 1.01 |
| ARF | 3 | 3583 | 3433 | 1016 | 7483 | 1.04 | 0.08 | 0.98 | 1.10 |
| ARS | 3 | 10,166 | 2819 | 6966 | 12,283 | 0.99 | 0.01 | 0.99 | 1.00 |

(Continued)
| Participant no. | Area     | No. of areas | Time mean | Time SD | Time min | Time max | CE mean | CE SD | CE min | CE max |
|----------------|----------|--------------|-----------|---------|----------|----------|---------|-------|--------|--------|
| 10 teller      | AMP      | 28           | 9305      | 9443    | 150      | 43,689   | 0.92    | 0.05  | 0.84   | 1.02   |
|                | AMR      | 13           | 2045      | 2104    | 533      | 8316     | 0.93    | 0.06  | 0.87   | 1.04   |
|                | ARF      | 30           | 3774      | 4450    | 233      | 17,633   | 0.99    | 0.06  | 0.86   | 1.09   |
|                | ARS      | 2            | 3299      | 1532    | 2216     | 4383     | 1.02    | 0.02  | 1.01   | 1.04   |
|                | AMPR     | 4            | 332       | 150     | 116      | 466      |         |       |        |        |
|                | ARMR     | 2            | 9058      | 10,241  | 1816     | 16,300   |         |       |        |        |
| 11 trans       | AMP      | 22           | 7302      | 9326    | 800      | 34,693   | 1.00    | 0.07  | 0.92   | 1.16   |
|                | AMR      | 6            | 941       | 774     | 233      | 2016     | 1.00    | 0.10  | 0.91   | 1.13   |
|                | ARF      | 20           | 4346      | 6178    | 216      | 25,066   | 1.02    | 0.08  | 0.90   | 1.27   |
|                | ARS      | 3            | 8433      | 561     | 7950     | 9050     | 0.99    | 0.02  | 0.96   | 1.00   |
|                | AMPR     | 1            | 883       |         |          |          |         |       |        |        |
|                | ARMR     | 1            | 4033      |         |          |          |         |       |        |        |
| 12 teller      | AMP      | 7            | 15,486    | 10,772  | 4850     | 34,492   | 1.02    | 0.04  | 0.96   | 1.05   |
|                | AMR      | 2            | 4458      | 5762    | 383      | 8533     | 0.95    |       |       |        |
|                | ARF      | 7            | 5490      | 4179    | 1666     | 11,916   | 0.99    | 0.08  | 0.90   | 1.10   |
|                | ARS      | 2            | 9349.50   | 5680    | 5333     | 13,366   | 1.01    | 0.10  | 0.94   | 1.08   |
|                | AMPR     | 3            | 1488      | 575     | 1100     | 2149     |         |       |        |        |
|                | ARMR     | 1            | 1250      |         |          |          |         |       |        |        |
| 13 teller      | AMP      | 19           | 10,881    | 11,642  | 1150     | 48,677   | 0.98    | 0.04  | 0.93   | 1.05   |
|                | AMR      | 6            | 2291      | 774     | 183      | 10,433   | 0.96    | 0.07  | 0.90   | 1.07   |
|                | ARF      | 20           | 5611      | 6032    | 616      | 26,800   | 1.00    | 0.04  | 0.94   | 1.10   |
|                | ARS      | 2            | 8358      | 4041    | 5500     | 11,216   | 0.99    | 0.07  | 0.94   | 1.03   |
|                | AMPR     | 1            | 2100      |         |          |          |         |       |        |        |
|                | ARMR     | 1            | 1300      |         |          |          |         |       |        |        |
| 14 trans       | AMP      | 10           | 19,485    | 14,870  | 983      | 49,910   | 1.06    | 0.04  | 1.00   | 1.10   |
|                | AMR      | 4            | 15,670    | 26,241  | 1116     | 55,000   | 1.01    | 0.09  | 0.95   | 1.08   |
|                | ARF      | 5            | 3016      | 4435    | 200      | 10,733   | 1.10    | 0.07  | 1.03   | 1.18   |
|                | ARS      | 1            | 6300      |         |          |          |         |       | 1.02   |        |
|                | AMPR     | 3            | 1160      | 900     | 483      | 2183     |         |       |        |        |
|                | ARMR     | 0            |           |         |          |          |         |       |        |        |
| 15 teller      | AMP      | 4            | 16,012    | 9076    | 5299     | 27,409   | 1.07    | 0.01  | 1.06   | 1.08   |
|                | AMR      | 2            | 3016      | 1673    | 1833     | 4200     | 1.04    | 0.04  | 1.01   | 1.07   |
|                | ARF      | 1            | 1533      |         |          |          |         |       | 1.00   |        |
|                | ARS      | 2            | 5091      | 3877    | 2350     | 7833     | 1.01    | 0.02  | 1.00   | 1.02   |
|                | AMPR     | 1            | 900       |         |          |          |         |       |        |        |
|                | ARMR     | 0            |           |         |          |          |         |       |        |        |

Average across subjects

|                | AMP      | 9.87        | 11,660.45 | 11,379.06 | 150      | 56,496   | 1.01    | 0.08  | 0.84   | 1.45   |
|                | AMR      | 4.67        | 5146.77   | 10,670.47 | 116      | 55,000   | 0.99    | 0.08  | 0.84   | 1.16   |
|                | ARF      | 8.6         | 4781.95   | 5326.00   | 200      | 26,800   | 1.02    | 0.07  | 0.86   | 1.27   |
|                | ARS      | 2.07        | 8889.48   | 4736.85   | 200      | 17,400   | 1.02    | 0.05  | 0.86   | 1.11   |
|                | AMPR     | 1.07        | 887.06    | 697.23    | 116      | 2183     | 1.02    | 0.06  | 0.95   | 1.15   |
|                | ARMR     | 1.07        | 2524.56   | 3867.50   | 233      | 16,300   | 1.02    | 0.08  | 0.91   | 1.14   |

trans: knowledge transformer; teller: knowledge teller; AMP: area of monitoring and production; AMR: area of monitoring and revision; ARF: area of review—formative; ARS: area of review—summative; ARMR: area of review, monitoring, and revision; AMPR: area of monitoring and production with review.

in the amount of time spent on the writing behaviors as indicated by the large standard deviations. For example, the shortest time for an AMP was 150 milliseconds and the longest was 56,496 milliseconds. At times, participants engaged in long and sustained production and review of their written texts and at other times rapidly transitioned through each writing behavior. In addition, there was a great deal of variability in the number of writing behaviors each participant...
used in the production of text. For instance, one participant had 28 sessions of text production (i.e. AMPs), whereas another participant had only two.

On average, across the 15 participants, the production of text, AMPs, represented the longest time (11,660.45 milliseconds). Review processes, ARFs and ARSs, represented about half as much time (4781.95 and 8899.48 milliseconds, respectively). Revision, AMRs, represented about as much time (5146.77 milliseconds) as ARFs and ARSs. Finally, combined processes, AMPRs and ARMRs, were rare and required little time (887.06 and 2524.56 milliseconds, respectively). Cognitive effort was, in general, fairly stable across the 15 participants. However, the maximum values indicate that the production of text (AMPs) and the review of text (ARF) were at times the most demanding. In sum, writing for this task was fast, involving quick transitions from one writing behavior to another, with time and cognitive effort being focused mainly on text production and review.

**Writing behaviors used in a problem-solving task**

Our final goal was to investigate how our writing methodology can be used to identify the writing behaviors students use in a problem-solving task. Our research question was, ‘Do the writing behaviors used by students who undergo a knowledge change differ from the writing behaviors used by students who do not experience a knowledge change?’ We identified the initial presentation of the informational/refutational text as the critical juncture for observing differences in writing behaviors. At this point in the study, participants had already produced an answer and then were provided with a text that either refuted or confirmed their answer. If participants gained new understanding about gravity from the text, they could revise their answers to reflect that new understanding. If participants either maintained their misconception of gravity or they already had a correct conceptualization of gravity, they could revise their answers but in ways that would not fundamentally change them. Therefore, we hypothesized that differences in writing behaviors would be evident depending on whether a participant changed his or her conceptualization of gravity or maintained their initial conceptualization of it.

Of the 15 participants, 9 underwent a change in their understanding of gravity and revised their answers. Of these, eight changed incorrect answers to correct answers, and one changed a correct answer to an incorrect answer. The remaining six participants showed no change in their understanding of gravity: Four of them retained incorrect answers, and two retained correct answers.

**Knowledge tellers versus knowledge transformers.** To characterize the two types of writers, we used Bereiter and Scardamalia’s (1987) knowledge telling and knowledge transforming models of writing. In the present case, we referred to participants who retained their answers as knowledge tellers and those who changed their answers to knowledge transformers. According to Bereiter and Scardamalia, knowledge tellers automatically retrieve existing linguistic and content knowledge and use that knowledge in their writing. There is little goal setting, reviewing, or planning of that writing. Knowledge telling is a natural and efficient approach to writing, and accordingly, the cognitive demands are on a par with the cognitive demands required by conversational discourse. In contrast, knowledge transformers approach writing with deliberate strategic control over all components of the writing process. This strategic control is cognitively demanding and can at times deplete cognitive resources. Rather than simply retrieving linguistic and content knowledge, knowledge transformers attempt to integrate the two to generate new knowledge. These changes in knowledge are considered and revised in accordance with evolving goals for writing. Based on these descriptions, we expected that knowledge transformers would allocate greater cognitive resources to all writing behaviors and engage in higher proportions of planning, reviewing, and revision than knowledge tellers.

**Data analysis.** We calculated the following descriptive statistics for each participant for periods before and after he or she accessed the informational/refutational text: the number of each writing behavior, the proportion of potential planning time within each behavior (i.e. the amount of additional time utilized by the participant to complete a specific writing behavior), the proportion of total time devoted to each behavior, and the average level of cognitive effort expended on each. Table 3 illustrates for participant 2 (the same participant illustrated in Table 1) that before receiving the informational/refutational text, he or she engaged in four AMPs, two ARFs, one ARS, one AMR, and zero AMPRs and ARMRs. With the exception of the last two behaviors, each of the writing behaviors included extra time compared to baseline so that there was additional time for potential planning to occur within each. For instance, 45% of the time devoted to AMPs was extra time. Across all of the writing behaviors for the initial answer, the participant allocated 45% of the time to writing the answer, 15% to reviewing the answer (13% as an ARF and 2% as an ARS), 2% to revising the answer, and 37% to potential planning of the answer, which was in addition to the extra time that was embedded in text production, review, and revision. Cognitive effort was at baseline for the ARS but was in excess of baseline for text production, review, and revision.

After reading the informational/refutational text, the participant revised his or her answer from incorrect to correct. The revision consisted of two AMPs, one ARF, one ARS, one AMR, and zero AMPRs and ARMRs. During revision, the participant changed his or her entire answer, which accounted for 40% of the time having been allocated to revising. In addition, the participant allocated 37% of the time to writing the new answer, 14% to reviewing the answer (2% as an ARF and 12% as an ARS), and 10% to potential planning of the
answer. Additional time was available both during text production (19%) and reviewing (40%). Cognitive effort was reduced to baseline or below baseline for each behavior except for reviewing, which showed increased effort and a large proportion of extra time spent on the review.

Because of the few number of participants in each of our two groups and the large amount of variability in the descriptive statistics, we restricted our analyses to a comparative description of the writing behaviors of knowledge tellers versus knowledge transformers. Table 4 shows the same kind of descriptive statistics that were described for the participant illustrated in Table 3, although the top panel in Table 4 contains descriptive statistics that were calculated by averaging across the six knowledge tellers, and the lower panel contains descriptive statistics that were calculated by averaging across the nine knowledge transformers. The knowledge tellers and knowledge transformers are identified in Table 2.
Before reading the informational/refutational text, both groups looked very similar in their writing behaviors. The number of each kind of behavior was similar as was the proportion of time allocated to each. About 50% of writing time was allocated to AMPs, and about 13–22% of the writing time was allocated to reviewing (i.e. ARFs and ARSs). Both groups allocated about 30% of their time to potential planning, and additional time was made available primarily during text production. Cognitive effort allocated to each writing behavior was similar between the two groups; however, knowledge transformers appear to have allocated more cognitive effort to revision than knowledge tellers (1.05 vs 0.95).

After reading the informational/refutational text, the groups’ writing behaviors looked much different. Although the proportion of time allocated to potential planning and text production was about the same for the two groups, the knowledge tellers produced nearly twice as many AMPs as knowledge transformers, and they allocated greater cognitive effort to them. The greater number of AMPs for knowledge tellers is reflected in the fact that they simply added additional words throughout their texts to elaborate on an already settled answer, whereas knowledge transformers deleted large portions of their answers and produced entire new sections of text.

There was an interaction between reviewing and revision between the two groups. Knowledge tellers devoted about 50% of their time to reviewing and only 4% of their time to revision. In contrast, knowledge transformers devoted only about 29% of their time to reviewing and 26% of their time to revision. Because knowledge tellers did not engage in wholesale revisions of their answers, their time was focused on reviewing what was already written. Knowledge transformers needed less time to review but more time to change an answer they knew needed to change in response to the information provided in the informational/refutational text.

A striking difference between the two groups is the allocation of cognitive effort to the various writing behaviors. Knowledge tellers allocated above baseline cognitive effort only to text production, but knowledge transformers were above baseline for text production, formative, and summative reviewing and for the combined process of reviewing text while they were deleting portions of an incorrect answer. In sum, knowledge transformers appear to have allocated more cognitive effort to changing their answers than knowledge tellers.

**Discussion.** When first presented with the problem about gravity, all students likely formed at least a partial answer while still reading the problem. When they navigated to the writing screen, they began to write that answer, devoting about 50% of their time to text production, 20% to reviewing their answers, and 30% of their time was additional time that might have been used to think about further developing that answer. All students spent little time revising their answers, which supports the notion that their answers were likely already in some form prior to writing. In general, were it not for the higher levels of cognitive effort allocated to almost all of the writing behaviors, the kind of writing demonstrated by all students initially had the characteristics of knowledge telling. Although keep in mind that the higher levels of cognitive effort were judged to be higher in relation to the writing behaviors required to write in response to the Jack and Jill problem. Therefore, the increased levels of cognitive effort for the gravity problem may have been in response to simply a more difficult task.

Differences in writing behaviors became evident after students read the informational/refutational text. All students allocated higher levels of cognitive effort and about 25% of their time to adding new text; however, the kind of text that was added differed. Knowledge tellers added text that mostly elaborated what was already written and did not change the answer in any fundamental way. In contrast, knowledge transformers, after deleting their answers, which was reflected in a higher proportion of revising, added text that changed their answers in fundamental ways, that is, from incorrect to correct or in one instance from correct to incorrect. This greater emphasis on revision by the knowledge transformers came at the price of less time allocated to reviewing in comparison to the knowledge tellers. One clear difference between the two groups was the overall allocation of cognitive effort to the writing behaviors. Adding text, whether wholesale changes or elaborations, demanded cognitive resources. However, for the knowledge transformers, the review of their wholesale changes was more cognitively demanding than for the knowledge tellers’ review of their elaborations.

Support for our predictions about knowledge tellers versus knowledge transformers was mixed. Contrary to our expectations, the proportion of time allocated to text production and planning was about the same for both groups, and although the knowledge transformers did allocate more time to revision, knowledge tellers allocated more time to reviewing. Both groups allocated higher levels of cognitive effort to text production, but as per our predictions, knowledge transformers allocated more cognitive effort to reviewing than knowledge tellers.

**General discussion.** We had three goals for the present research: (a) review the literature on existing writing research methodologies, (b) describe our new writing methodology, and (c) conduct a pilot study that demonstrated how our methodology could be used to investigate the use of writing in problem solving. In our review of existing writing research methodologies, we noted the strengths and weaknesses of each and described how our methodology could improve on each. Like existing writing research methodology, the methodology that we developed has the power to examine the larger patterns and fluctuations of cognitive processes involved in writing.
However, our methodology also provides the fine-grained detail necessary for understanding the complex dynamics of those processes. There are still issues to be resolved in our methodology, but we believe that future iterations may be able to address them.

Although not problem free, our new methodology provides several important advantages over existing methodologies. TRAKTEXT provides continuous time-course measures for all writing behaviors that are observed. Some of these behaviors can be as rapid as 116 milliseconds and as long as 56,496 milliseconds, but most average between 5 and 15 seconds. In addition, movement from one writing behavior to the next can be rapid. For example, in the data presented in Table 1, the writer went from text production to review to production to review and then to revision in 85 seconds. Such rapidity and changeability from one behavior to another could never be captured by a verbal-protocol analysis. In addition, rather than randomly sampling for measures of cognitive effort, TRAKTEXT also captures continuous measures of pupil dilation that can be used to calculate measures of cognitive effort for each behavior. Importantly, capturing cognitive effort with eye tracking does not interrupt the writer from the writing task, which is a shortcoming of using secondary RTs.

TRAKTEXT also allows the direct observation of writing behaviors. Rather than asking writers to identify the writing behaviors they are currently using by selecting from a list of a prior writing behaviors, writers can simply engage in a natural course of writing without interruption. The behaviors are then identified by direct analysis of the data. This approach allows for the identification of multiple processes, some of which can occur simultaneously, such as reviewing and revision or reviewing and text production. Direct observation of writing behaviors also eliminates forcing the writer into theory-driven decisions about the kinds of behaviors being used, which allows for the expansion and reconceptualization of writing theory.

The elimination of verbal reports from the analysis of writing is also a benefit provided by TRAKTEXT. There is no need for potential contamination of cognitive processes imposed by simultaneously reporting on and engaging in a writing process. The writing process currently being reported is not confounded by the writer’s interpretation of that process. There is no need to choose between reports of two or more parallel processes; therefore, less salient processes are not compromised by more dominant ones. And the cumbersome analysis of verbal data is eliminated.

TRAKTEXT provides many advantages over existing writing research methodologies, but there are at least four limitations we have identified that need to be addressed. First, establishing a more reliable base rate for both key-boarding speed and cognitive effort is needed. We established both base rates by asking participants to engage in a simple writing exercise. The writing samples were short and did not provide for greater variability in typing speeds nor cognitive effort. In future work, requiring a longer writing sample would help with calculating more reliable measures of both.

Second, we calculated potential planning by measuring extra time spent on a writing behavior as compared to the base rate. What writers actually did during those periods of extra time is unknown. The extra time could have been for planning or daydreaming. Because of the rapidity of the writing behaviors, participants likely were not engaging in much of the latter and were likely more engaged in the former; however, a definitive answer to this is not known.

Third, the kinds of writing behaviors we did observe may be unique to the type of writing task we asked participants to respond to. The purposes for writing can have strong and fundamental impact on writing. A different task may have required different behaviors. Future research needs to examine writing under other conditions before the categories of writing we observed can be validated.

Fourth, in our pilot study to examine writing behaviors, we had intended to manipulate writing using a problem-solving paradigm that has been shown to lead to knowledge change. Our assumption was that writing would differ between participants who experienced a knowledge change versus those who did not. We did find differences that could be explained using the theoretical work concerning knowledge telling versus knowledge transformation. However, what role writing played in knowledge change is obscure. Participants could have changed their knowledge during reading and then simply recorded that change in writing. Alternatively, participants could have used writing to stimulate that change. If the latter, future research will need to focus more on strictly writing tasks without confounding writing with reading of supplied texts.

TRAKTEXT is a new methodology for the investigation of writing that has provided some new insights into writing, but it is still at the beginning stages of development. There are issues that remain to be addressed. Our initial results have demonstrated that writing appears to be nearly a frenetic process, proceeding in the “spurts and pauses” described by Kellogg (1987a: 273). Writing behaviors can occur uninterrupted over longer periods of time and require low levels of cognitive effort but at other times can rapidly change from revision to text production to review and back to revision, all requiring high levels of cognitive effort. Every writing behavior that we identified has the potential to change to any other writing behavior in time periods of less than a second. Although these changes seemingly occur in a chaotic fashion, the end result is the production of written text that conveys meaning. Writing is the production of meaning, but the production can occur in a myriad of ways. More work needs to be done, but we believe that the potential for future uses of our technology is considerable.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

Note
1. The TRAKTEXT software and associated software will be made available to interested parties upon request to the first author.

References
Ahern S and Beatty J (1979) Pupillary responses during information processing vary with scholastic aptitude test scores. Science 205: 1289–1292.
Alamargot D, Chesnet D, Dansac C, et al. (2006) Eye and pen: A new device for studying reading during writing. Behavior Research Methods 38: 287–299.
Alamargot D, Dansac C, Chesnet D, et al. (2007) Parallel processing before and after pauses: A combined analysis of graphomotor and eye movements during procedural text production. In: Torrance M, van Waes L and Galbraith D (eds) Writing and Cognition: Research and Applications. New York: Elsevier Science, pp. 13–29.
Alamargot D, Leuwers C, Caporossi G, et al. (2012) Eye-tracking data during written recall: Clues to subject-verb agreement processing during translation. In: Berninger VW (ed.) Past, Present, and Future Contributions of Cognitive Writing Research to Cognitive Psychology. New York: Psychology Press, pp. 441–459.
Beatty J (1982) Task-evoked papillary responses, processing load, and the structure of processing resources. Psychological Bulletin 91: 276–292.
Beers S (2004) Reading fluency and adolescent students’ reading processes during writing. Unpublished Doctoral Dissertation.
Bereiter C and Scardamalia M (1987) The Psychology of Written Composition. Hillsdale, NJ: Lawrence Erlbaum.
Boland JE (2004) Linking eye movements to sentence comprehension in reading and listening. In: Carreiras M and Clifton C Jr (eds) The On-line Study of Sentence Comprehension: Eyetracking, ERPs and Beyond. New York: Psychological Press, pp. 51–76.
Carreiras M and Clifton C Jr (2004) On the on-line study of language comprehension. In: Carreiras M and Clifton C Jr (eds) The On-line Study of Sentence Comprehension: Eyetracking, ERPs and Beyond. New York: Psychological Press, pp. 1–14.
Ericsson KA and Simon HA (1980) Verbal reports as data. Psychological Review 87: 215–251.
Eye Tracker System Manual, ASL EyeTrac 6 (2007) Bedford, MA: Applied Science Laboratories, An Applied Science Group Company.
Fitzgerald J (1987) Research on revision in writing. Review of Educational Research 57: 481–506.
Granholm E, Asarnow RF, Sarkin AJ, et al. (1996) Pupillary responses index cognitive resource limitations. Psychophysiology 33: 457–461.
Guzzetti BJ, Snyder TE, Glass GV, et al. (1993) Promoting conceptual change in science: A comparative meta-analysis of instructional interventions from reading education and science education. Reading Research Quarterly 28: 117–159.
Hayes JR and Flower LS (1980) Identifying the organization of writing processes. In: Gregg LW and Steinberg ER (eds) Cognitive Processes in Writing. Hillsdale, NJ: Lawrence Erlbaum, pp. 3–30.
Hayes JR and Nash JG (1996) On the nature of planning in writing. In: Levy CM and Ransdell S (eds) The Science of Writing: Theories, Methods, Individual Differences, and Applications. Mahwah, NJ: Lawrence Erlbaum, pp. 29–55.
Hynd CR and Guzzetti B (1998) When knowledge contradicts intuition: Conceptual change. In: Hynd CR (ed.) Learning from Text Across Conceptual Domains. Mahwah, NJ: Lawrence Erlbaum, pp. 139–163.
Janssen D, van Waes L and van den Bergh (1996) Effects of thinking aloud on writing processes. In: Levy CM and Ransdell S (eds) The science of writing: Theories, methods, individual differences, and applications. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 230–250.
Just MA and Carpenter PA (1993) The intensity dimension of thought: Pupillometric indices of sentence processing. Canadian Journal of Experimental Psychology 47: 310–339.
Kahneman D (1973) Attention and Effort. Englewood Cliffs, NJ: Prentice-Hall.
Kahneman D and Beatty J (1966) Pupil diameter and load on memory. Science 154: 1583–1585.
Kellogg RT (1986) Designing idea processors for document composition. Behavior Research Methods, Instruments, & Computers 18: 118–128.
Kellogg RT (1987a) Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. Memory & Cognition 15: 256–266.
Kellogg RT (1987b) Writing performance: Effect of cognitive strategies. Written Communication 4: 269–298.
Kellogg RT (1988) Attentional overload and writing performance: Effects of rough draft and outline strategies. Journal of Experimental Psychology: Learning, Memory, and Cognition 14: 355–365.
Levy CM and Ransdell S (1994) Computer-aided protocol analysis of writing processes. Behavior Research Methods, Instruments, & Computers 26: 219–223.
Levy CM and Ransdell S (1996) Writing signatures. In: Levy CM and Ransdell S (eds) The Science of Writing: Theories, Methods, Individual Differences, and Applications. Mahwah, NJ: Lawrence Erlbaum, pp. 149–161.
McConkie GW, Hogaboam TW, Wolverton GS, et al. (1979) Toward the use of eye movements in the study of language processing. Discourse Processes 2: 157–177.
Mitchell DC (2004) On-line methods in language processing: Introduction and historical review. In: Carreiras M and Clifton C Jr (eds) The On-line Study of Sentence Comprehension: Eyetracking, ERPs and Beyond. New York: Psychological Press, pp. 15–32.
Nisbett RE and Wilson TD (1977) Telling more than we can know: Verbal reports on mental processes. Psychological Review 84: 231–259.
Pickering MJ, Frisson S, McElree B, et al. (2004) Eye movements and semantic composition. In: Carreiras M and Clifton C Jr (eds) The On-line Study of Sentence Comprehension: Eyetracking, ERPs and Beyond. New York: Psychology Press, pp. 33–50.
Piolat A, Olive T, Roussey J-Y, et al. (1999) Scriptkell: A tool for measuring cognitive effort and time processing in writing and other complex cognitive activities. Behavior Research Methods, Instruments, & Computers 31: 113–121.
Postma A (2000) Detection of errors during speech production: A review of speech monitoring models. Cognition 77: 97–131.
Rayner K (1998) Eye movements in reading and information processing: 20 years of research. Psychological Bulletin 124: 372–422.
Smallwood J and Schooler JW (2006) The restless mind. *Psychological Bulletin* 132: 946–958.

**Author biographies**

**Dr. Douglas J. Hacker** is a full professor in the Department of Educational Psychology. He received his Ph.D. in educational psychology from the University of Washington in 1994. From 1994 to 1999, Dr. Hacker was an assistant/associate professor in the Department of Counseling, Educational Psychology and Research at The University of Memphis. During those years, he worked in the areas of reading and writing processes, metacognition, self-regulated learning, teacher education, and school and program evaluation. Dr. Hacker moved to the University of Utah in 1999 and has continued his research in the previous areas and has added to them research in the area of the detection of deception.

**Dr. Matt C. Keener** is an associate instructor in the Department of Educational Psychology at the University of Utah. He received his Ph.D. in educational psychology from the University of Utah in 2011. From 2014 to 2015, he also taught as an adjunct instructor at Salt Lake Community college. His research interests include the cognitive processes involved in reading, writing, and metacognition.

**Dr. John C. Kircher** earned his B.S. from The Pennsylvania State University and his M.S. and Ph.D. in Psychology from the University of Utah. He has been a professor of psychology and educational psychology at the University of Utah since 1986. In 1990, he and a colleague developed the first computerized polygraph system that included electronic circuitry for recording physiological data, computer algorithms to extract diagnostic features from multiple physiological measures, and Bayesian methods to assess the probability that the examinee was deceptive on the test. In 2006, Dr. Kircher and his colleagues developed the Ocular-motor Deception Test that detects deception from changes in pupil size, response time, and ocular-motor reading measures.