ascilite

Tracking reading strategy utilisation through pupillometry

Aaron Y. Wong and **Jarrod Moss**
Mississippi State University

Christian D. Schunn
University of Pittsburgh

Explicit reading strategies help low-knowledge readers make the inferences necessary to comprehend expository texts. Self-explanation is a particularly effective strategy, but it is challenging to monitor how well a reader is applying self-explanation without requiring the reader to externalise the self-explanations being generated. Studies have shown that different reading strategies vary in the amount of cognitive control required as well as the engagement of brain regions involved in internally-directed attention. Pupil diameter is related to task engagement and cognitive control via the brain's locus coeruleus-norepinephrine system. Therefore, pupil diameter could be a method to unobtrusively measure a reader's use of self-explanation. The current study assessed whether pupil diameter can be used to distinguish between the use of different reading strategies and whether it is linked to the quality and effectiveness of the strategy in terms of learning gains. Participants reread, paraphrased, and self-explained texts while pupil diameter was recorded, and completed comprehension tests. Average pupil diameter differed between all three reading strategies, and pupil diameter was related to learning gains and the quality of strategy use. The results suggest that pupil diameter could be used to track effective reading strategy utilisation.

Introduction

One of the problems that students face when learning or studying with textbooks is that textbooks are written in a way that requires extensive inferences to be made in order to comprehend the texts, especially in ways that lead to knowledge revision. However, readers with low domain knowledge have great difficulty making these inferences (McNamara & Kintsch, 1996; McNamara, Kintsch, Songer, & Kintsch, 1996). Reading strategies can be used to facilitate making some of the necessary inferences (McNamara, 2004), yet not all reading strategies are equally effective, and studies have shown that differences exist in the strategies used by good and poor readers (Chi, Bassock, Lewis, Reimann, & Glaser, 1989; Just & Carpenter 1992). One of the more common reading strategies used by students is rereading a text (Carrier, 2003; Karpicke, Butler, & Roediger, 2009). However, the benefits of rereading are quite limited and rereading has been shown to be less effective than other reading strategies (Callendar & McDaniel, 2009; Chi, De Leeuw, Chiu, & LaVancher, 1994; McDaniel, Roediger, & McDermott, 2007; Moss, Schunn, Schneider, & McNamara, 2013; Moss, Schunn, Schneider, McNamara, & VanLehn, 2011). Students can be trained in more effective reading strategies such as self-explanation, but it takes extensive practice to be able to robustly and automatically use such strategies. Self-explanation is a reading strategy in which a student explains the text to himself or herself in a variety of ways as if explaining it to another person. Unfortunately, it is challenging for students, teachers, and researchers to monitor students' use of self-explanation. Therefore, creating new methods and systems for monitoring use of self-explanation could benefit students. Here we propose a new method based on pupil dilation.

Self-explanation

Studies have shown that the use of self-explanation can lead to improvements in test performance and deeper understanding of texts (Chi et al., 1989, 1994). Self-explanation is thought to improve comprehension because gaps in the learner's representation of the text are identified as the learner tries to explain the text content. Inference processes that improve the coherence of the text representation by filling in those gaps can then be used. Despite the long-established and highly cited benefits of self-explanation, many readers may not use self-explanation at all or are unable to perform the reading strategy adequately (Chi et al., 1989; McNamara, 2004). In order to improve the use of self-explanation, McNamara (2004) developed an instructional method called self-explanation reading training (SERT).

SERT aims to provide instruction on how to use self-explanation. SERT divides self-explanation into five techniques: comprehension monitoring, paraphrasing, bridging, elaboration, and prediction. Comprehension monitoring involves the student being aware of his or her own understanding of what is being read. As the explanation of the text is being constructed, points where necessary information is missing from memory should be identified via monitoring and then the other SERT strategies can help to fill in that missing knowledge. Paraphrasing involves the student restating what is being read in his or her own words. Paraphrasing helps to activate relevant knowledge in memory that will help in making inferences as the student tries to rephrase the text in his or her own words. Bridging is making inferences between a current text and what the student has read before. Elaboration is integrating prior knowledge with the new information being provided in a text. These two inference processes are responsible for helping to fill in the identified gaps in knowledge identified via monitoring. Prediction is trying to anticipate future information in the text. The use of these five techniques helps a reader learn from textbooks by aiding in the process of comprehension.

SERT has been implemented into an automated reading strategy tutor called the Interactive Strategy Trainer for Active Reading and Thinking (iSTART) (McNamara, Levinstein, & Boonthum, 2004; McNamara, O'Reilly, Best, & Ozuru, 2006). iSTART is a web-based intelligent tutoring system that teaches students how to perform the five SERT techniques using natural language processing (i.e., computer-based processing of students' typed explanations) and animated agents. The use of an automated version of SERT has several benefits as it does not require an observer to be trained on how to use SERT or be present during the sessions to provide feedback. Students who have used SERT and iSTART have shown improvements in test performance and a deeper understanding of expository material (McNamara, 2004; McNamara et al., 2004, 2006).

A problem that iSTART and other automated reading strategy tutors face is whether the tutor can be applied to a more general setting. For example, one limitation of iSTART is that lessons are designed around specific texts. In order for students to receive useful feedback while self-explaining a text, iSTART must be programmed with the necessary information on what techniques are appropriate. In addition, automated reading strategy tutors require users to type explanations in a user interface in order for the strategies to be monitored (Graesser, McNamara, & VanLehn, 2005). Once a reader is no longer using this typed-response computer interface, it cannot be determined if the reader is still using the techniques taught by the tutor (e.g., while independently reading a non-electronic textbook). In addition, students may find typing self-explanations too slow and intrusive. One potential solution to these problems is to develop an alternative method for determining what reading strategy is being performed that is not intrusive and can be applied to any text. Once the reading strategy performed is determined, feedback can then be given on how to improve performance of the strategy if necessary.

Past studies using neuroimaging have shown that the level of cognitive control tends to vary depending on the characteristics of a reading strategy (Moss et al., 2013, 2011). For example, Moss and colleagues (2013, 2011) found that rereading a text does not engage cognitive control regions of the brain as much as the more effective self-explanation strategy. The researchers contrasted brain activity in rereading, self-explanation, and the paraphrasing SERT technique without use of the other SERT techniques. Cognitive control activity was significantly different for rereading and the other two strategies, but self-explanation and paraphrasing did not differentially activate cognitive control regions. However, a number of brain regions involved in internally-directed attention, memory retrieval, and inferencing were found to be more active for self-explanation than for the other two strategies (e.g., posterior cingulate cortex, dorsomedial prefrontal cortex, and inferior parietal cortex). These regions have been hypothesised to be a coherence-building network that is used during discourse comprehension to create coherent (i.e., well connected) internal representations of the material read (Moss & Schunn, 2015). Therefore, it should be possible to monitor the brain to determine whether students are utilising effective reading strategies. However, this past research involved participants performing the reading strategies inside a magnetic resonance imaging scanner, which is not a normal reading environment and prohibitively expensive to use in any applied context or research involving large numbers of participants. At the same time, these results do indicate that monitoring the degree of cognitive control and other cognitive activity should be a feasible way to monitor reading strategy use if a more accessible technology could be used.

Pupil diameter, cognitive control, and strategic reading

One method that has been used to study comprehension processes during reading is to use various forms of eye measures (Rayner, Chace, Slattery, & Ashby, 2006). Eye measures, in particular pupil dilation, have been shown to be able to measure cognitive load during the performance of cognitive tasks (Beatty, 1982; Peysakhovich, Causse, Scannella, & Dehais, 2015; Querino et al., 2015). This relationship is caused by the relationship between pupil diameter and the locus coeruleus-norepinephrine system (LC-NE) in the brain. While an exact anatomical pathway for the influence of activity in the LC-NE system on pupil diameter has yet to be clearly identified, there is a clear relationship between LC-NE activity and pupil diameter across a range of methodologies and tasks (Aston-Jones & Cohen, 2005; Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Murphy, Robertson, Balsters, & O'Connell, 2011). The LC-NE system is thought to be related to engagement in goal or task-related processes, and therefore pupil diameter is a marker of the amount of task-relevant cognitive processing (Aston-Jones & Cohen, 2005).

With recent developments in technology, eye movements can be recorded with wearable glasses. Furthermore, there are a number of websites describing methods for building low-cost eye trackers (e.g., $50-$100) that can measure pupil diameter. These apparatus are both low-cost and more similar to normal reading environments than a magnetic resonance imaging scanner. Because of the relationship between pupil dilation and cognitive processing, it should be possible to observe the cognitive load differences between reading strategies using an eye tracker that can measure pupil diameter. However, it is not clear whether differences between strategies that exhibit similar cognitive control processing in the brain will be differentiable based on pupil diameter alone. Therefore, the current study examines whether it is possible to detect the difference between reading strategies including rereading, paraphrasing, and self-explanation using pupil diameter.

Current study

In the study, participants were initially taught how to self-explain using the SERT techniques. Participants then reread, paraphrased, and self-explained texts while eye measures were recorded. Greater pupil diameter was predicted for reading strategies that require more cognitive processing that is, pupil diameter should be greater for texts that are self-explained or paraphrased than reread. If large and consistent differences in pupil diameter are observed between the reading strategies, it would suggest that pupil diameter might be used to monitor reading strategy use using less expensive equipment and in non-laboratory settings.

If pupil diameter does discriminate amongst reading strategies, the next question is whether the difference in cognitive processing between strategies as measured by pupil diameter is related to the difference in learning gains. Studies have shown that self-explanation benefits learning, and therefore the more that this strategy is used the greater the learning gains should be. This pupil-diameter-to-learning correlation would provide further support that the pupil diameter differences between strategies are related to comprehension processes that benefit learning rather than some other side effect of using a particular strategy that might be less important for learning.

If a positive pupil-diameter-to-learning correlation is found, then another question which could be examined is whether pupil diameter will provide more fine-grained information about the degree to which an individual is successfully using a reading strategy (i.e., variation within strategy use rather than only variation in which strategy is used). Past research has found individual differences in the degree to which individuals use self-explanation effectively (Chi et al., 1994). For example, could the quality of a self-explanation be determined through an analysis of pupil diameter? Here we define self-explanation quality as the amount of comprehension-aiding inferences as compared to other statements (e.g., paraphrases) that are made while self-explaining. It is known that greater self-explanation quality construed in this way is associated with greater learning gains (McNamara et al., 2004, 2006). A positive relationship between pupil diameter and strategy quality would suggest that pupil diameter would be able to discriminate between good and poor self-explanations and therefore provide more detailed information beyond whether self-explanation was being used.

Method

Participants

Twenty-two native English speakers were recruited from the University of Pittsburgh and Carnegie Mellon University communities (13 female, *M* age = 19.7 years, *SD* = 1.5, range = 18-24). None of the participants had taken a college physics course. All participants were paid for their participation.

Materials

Reading materials for the experiment consisted of three texts that were each about one introductory physics topic: DC circuits, pulley systems, and classical mechanics (i.e., forces and motions of objects). Each of the three texts consisted of 15, 2-4 sentence paragraphs. Each text also contained either 13 or 14 diagrams such that almost every paragraph had one accompanying diagram. Text comprehension questions for each text consisted of 15 multiple-choice questions. Pilot norming studies were performed with a separate group of 34 participants to ensure that all three texts and tests had equal difficulty.

Design

The study used a repeated measures design contrasting the three reading strategies: rereading, paraphrasing, and self-explanation. For a given participant, each text was assigned to one of the three reading strategies with the assignment of the texts to reading strategy counterbalanced across participants. Therefore, each participant always applied the same reading strategy for one entire text. The presentation order of the texts was randomised across participants.

Procedure

The study consisted of two sessions (pretest and strategy training, followed by eye-tracking and posttest), separated by 2-5 days. Eye tracking data were collected only during the second session.

Session 1
At the beginning of the first session, participants were given 30 minutes to complete a pretest. The pretest consisted of all the test questions for each of the three texts (45 in total). Participants then completed a training session on how to perform self-explanation.

The experimenter explained to the participant that he/she would be going through a tutorial on self-explanation and that these techniques would be used in the second session. Participants were shown a slideshow with audio that contained examples of someone performing the five SERT techniques: comprehension monitoring, paraphrasing, elaboration, bridging, and prediction. After a definition of each technique, participants were asked to identify which technique was in use during each example. After the examples, participants were given practice in using each technique. Participants first tried examples using single techniques and then moved on to doing full self-explanations with all of the techniques at once. For the single technique stage of the practice, the experimenter provided feedback on the practice attempts, and participants were required to repeat an attempt to use a technique until they successfully used it. For the use of all-techniques-at-once part of the practice, the experimenter required individuals who used fewer than two of the techniques on a given passage to repeat the practice on that passage using additional SERT techniques. The training session lasted approximately 20-30 minutes. Because all participants already understood rereading and paraphrasing was taught as part of SERT, no additional strategy training was needed for these other two strategies.

Session 2
The second session occurred 2-5 days after the first session (varying by participant availability, breaks for the weekend, and open time slots) in order to reduce the chance that participants would read the passages with the pretest questions in mind. Participants were first given a review of the self-explanation techniques. In this review, participants were shown an example of a good self-explanation of a paragraph with the individual self-explanation techniques highlighted in a colour-coded scheme allowing each statement to be matched to one of the SERT techniques. Participants were then given an opportunity to ask the experimenter any questions. The review lasted about 15 minutes.

Then the participants began the task of reading and using the reading strategies on the three physics text, with eye measures being collected. Participants were seated in front of the eye tracker and given instructions and training on how to complete the experimental tasks using the three reading strategies. During training, participants saw paragraphs from a practice text that was of a similar expository nature, but contained different content than the texts in the experiment.

The paraphrasing strategy was one of the techniques introduced during the self-explanation training, and participants were told to paraphrase out loud each sentence in the text by putting it into their own words without using any of the other self-explanation strategies. For the rereading strategy, participants were told to keep reading the text out loud until the computer indicated it was time to move to the next set of texts. Participants were prompted to stop rereading and move on by the instructions prompt at the bottom of the screen starting to flash. The amount of time allotted for rereading was 38 seconds to roughly equate the amount of time spent rereading with the amount of time spent paraphrasing and self-explaining; the time was determined from a pilot study involving the three strategies applied to the same texts.

Each of the three texts was divided into three blocks of five paragraphs each. The first block of each text was presented before the second block and so on. In other words, each block of a text was separated by a block from the other two texts (i.e., Text1-Section1, Text2-Section1, Text3-Section1, Text1-Section2, Text2-Section2, Text3-Section2, Text1-Section3, Text2-Section3, Text3-Section3). The presentation of the blocks was ordered in this manner to help control for confounding reading strategy with temporal effects such as fatigue or motivational change. Although a little artificial, this jumping around across topics during studying from text is not completely dissimilar from a student reviewing a textbook for an exam.

Before each block of five paragraphs, participants read instructions on the screen indicating the learning strategy they were to use for that block. These instructions also provided the title of the text. For each paragraph, participants first read the text without using any reading strategy. Participants pressed a keyboard button when finished reading. After the initial reading, a reminder of what reading strategy to use was shown at the bottom of the screen and the paragraph was displayed again so that the designated reading strategy could be used. Following the performance of the strategy, participants answered a question to assess the degree of mind wandering on the previous paragraph. They were asked how often in the previous paragraph they caught themselves thinking about topics other than the text they were reading. They provided a response on a scale from 1-7 with 1 being *not at all* and 7 being *frequently*. The process was then repeated for the remaining paragraphs in a block. This procedure was repeated for each block of each text. Participants were allowed to take a short break between each block. Reading of the texts and using the reading strategies lasted about 1 hour.

When reading or using a reading strategy, the title of the text was centered at the top of the screen. Text paragraphs were centered on the left half of the screen and accompanying diagrams were centered on the right half of the screen. Centered at the bottom was the reminder for which reading strategy to use. After all the sections of the text had been read and the reading strategies were used, participants completed the posttest. The posttest was the same as the pretest and participants were given 30 minutes to complete the posttest.

Eye data acquisition and processing

Eye measurements were collected using a Tobii 1750 eye tracker. The eye tracker is capable of capturing both eye movement and pupillometry measures. It is integrated into a 17-inch TFT monitor. The eye-tracker sampled eye-data (fixation location and pupil diameter) using binocular tracking at a rate of 50 Hz with an accuracy of 0.5–0.7 degrees. The Tobii 1750 does not require head stabilisation and allows for freedom of head-movement (30 x 15 x 20 cm at 60 cm from tracker). As a result, it is not one of the highly precise eye-trackers used by basic visual or reading researchers, and more closely resembles the capabilities of recently developed inexpensive eye-trackers.

Eye data collection was performed in a room with indoor lighting and no windows. Participants wearing glasses can sometimes pose a problem in calibrating the eye tracker, but the Tobii eye tracker usually performs well even with glasses. Tobii's calibration routine with nine calibration locations was used. The

calibration did not fail for any of the participants in this study. Before analyses of pupil diameter were performed, eye measurements during blinks and off-screen fixations were filtered out of the data. In addition, eye measurements in which at least one pupil was not detected were filtered out of the data.

Results

Analyses validating strategy effects on learning

Table 1 shows the average pre/posttest learning gains for each strategy. Learning gains were calculated as the percentage of maximum improvement using the following equations, where post is the percentage correct on the posttest and pre is the percentage correct on the pretest (Cohen, Cohen, Aiken, & West, 1999; Marx & Cummings, 2007). Learning gain was calculated using the equation (post-pre)/(100-pre). However, in the rare cases in which posttest performance was equal to or less than pretest performance, learning gain was calculated using the modified equation (post-pre)/(pre).

Table 1
Average learning gains by reading strategy condition

Reading Strategy	*M*	*SD*
Rereading	0.27	0.26
Paraphrase	0.35	0.30
Self-Explain	0.43	0.22

Because of prior research showing that self-explanation is better than paraphrasing alone and that paraphrasing is sometimes better than rereading (Moss et al., 2013, 2011), learning gains were analysed in a set of three planned pairwise strategy contrasts using paired t-tests. Self-explanation learning was greater than that for rereading, $t(21) = 2.19$, $p = .040$, $d = .47$. No other comparisons were significant. The lack of other significant differences is likely due to the small number of participants in this study. The means are ordered as predicted, and the primary purpose of this study was to examine whether pupil diameter could discriminate between the three reading strategies. The goal of the study was not to replicate the effect of reading strategy differences on learning gains that has been found in prior research using the same paradigm (Moss et al., 2011) in addition to other studies showing the benefits of self-explanation over other reading strategies (McNamara, 2004).

Can reading strategies be detected by pupil diameter?

To address this first research question, pupil diameter during strategy performance was examined. Pupil diameter was calculated by taking the average of both pupils and measured in millimetres. To create equal time intervals for each condition, eye measurements for the first 38 seconds of strategy use on a slide were analysed in 10, 3.8 second intervals. A linear mixed-effects model was used to assess the effect of strategy and interval on average pupil diameter. Using this analysis, crossed random effects due to both participant and text could be modeled in the same analysis without the assumption of sphericity inherent in ANOVA. These analyses were conducted using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) for *R*, and degrees of freedom and *p* values were calculated using Sattherwaite approximations as implemented in the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2013). The model selection approach used is that recommended by Barr, Levy, Scheepers, and Tily (2013) in which the maximal random effects structure supported by the data was used and then reduction of fixed effects started with a fully loaded model and proceeded by a backwards stepwise procedure.

The fully loaded model used strategy as a discrete factor variable and interval as a continuous variable as fixed effects and included a quadratic term for interval. The model also included random intercepts for text and subject and random slopes for strategy and interval for subject. Figure 1 shows the effects of reading strategy and interval on average pupil diameter. There was a significant interaction between strategy and the quadratic interval term, $t(8,958) = 9.27$, $p < .001$. The interaction appears to be driven mainly by the difference in the curve of the reread condition from the curves of the paraphrase and self-explanation condition. Interestingly, rereading showed a more rapid decline in pupil diameter from interval 2 to interval 4, whereas both paraphrasing and self-explanation showed a slow decline after

interval 2 through to interval 7. The increase in pupil diameter for rereading during the last interval is likely due to the prompt to stop rereading and move on. As a reminder, rereading was timed because unlike self-explanation or paraphrasing rereading had no subjective endpoint.
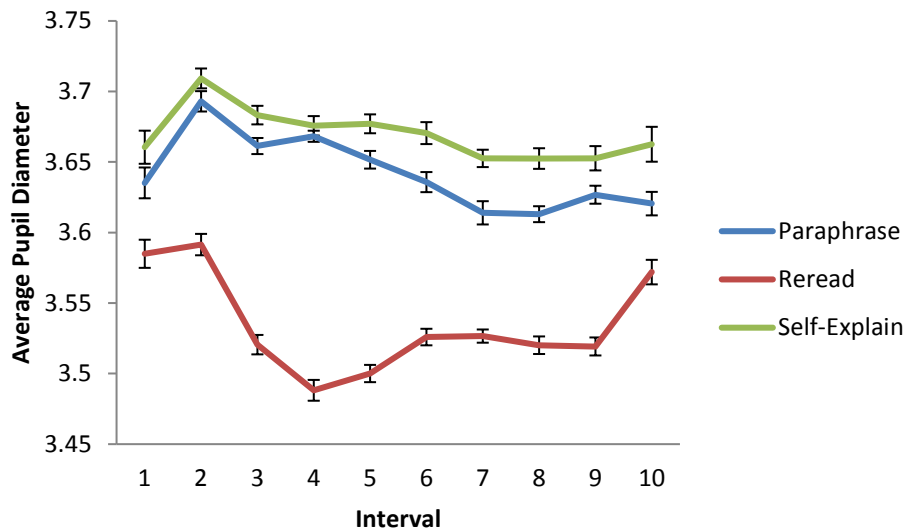

Figure 1. Average pupil diameter by reading strategy and interval. Error bars represent within-subject standard error (Loftus & Masson, 1994).

The fully loaded model was highly complex. It would be difficult to systematically unpack all of the effects, and many of the effects are not of theoretical interest (e.g., interactions at different time points with different texts). Therefore, simpler models were constructed to be able to more directly answer the questions of interest. Because the reread condition displayed different behaviour than the other two conditions and the main question of interest has to do with differences between self-explanation and paraphrasing, all subsequent analyses focused on examining the differences between the paraphrase and self-explanation conditions. A linear mixed-effects model was used to assess the effect of strategy and interval on average pupil diameter. The model included fixed effects for strategy and interval, random intercepts for text and subject, and random slopes for strategy and interval within subjects. Table 2 shows the results of the LME analysis. Average pupil diameter was greater for the self-explain condition than the paraphrase condition. There was also an effect of interval, and the interaction between strategy and interval only approached significance.

The differences in levels of cognitive processing can be observed during certain time periods, with a more rapid and immediate rise in pupil diameter combined with a slower decline in pupil diameter for the paraphrase and self-explanation strategies. In sum, the differences in average pupil diameter suggest that different levels of cognitive processing are required to perform the three reading strategies and that pupillometry can detect these differences.

Table 2
LME fixed effects results of strategy and interval on pupil diameter

	Estimate	*SE*	*df*	*p* value
Strategy	0.026	0.012	20.04	0.042
Interval	-0.020	0.004	22.03	<.001
Strategy x Interval	0.001	0.004	29.52	0.071

Note. Paraphrasing was set as the intercept for strategy comparison so that the positive estimate indicates a larger pupil diameter for self-explanation over paraphrasing.

Is pupil diameter predictive of differences in learning gains?

Because differences in pupil diameter were observed among the three strategies, analyses were performed to address the second research question of determining whether pupil diameter at various time points in strategy use were related to differences in learning. A linear mixed model with pupil diameter and strategy as fixed effects and including random intercepts for subject and text and a random slope for strategy within text was run for each interval. Because there is only one learning gain per condition, it is not possible to include interval as a fixed effect (i.e., all intervals for a given participant would have the same learning gain).

An interaction term was included in this model, but the interaction of strategy and pupil diameter was not significant at any interval. Strategy was also never a significant predictor. Table 3 shows the fixed effects result of average pupil diameter on learning gains at each time interval. There was a significant effect of pupil diameter on learning gains in intervals 4-6, 8, and 10. These results indicate that pupil diameter was more predictive of learning gains than simply knowing which strategy participants were using.

Table 3
LME fixed effect results of average pupil diameter on learning gains by interval

Interval	Estimate	SE	df	t value	p value
1	0.038	0.131	31.37	0.289	0.774
2	0.003	0.195	38.57	0.019	0.985
3	0.074	0.157	29.22	0.472	0.641
4	0.381	0.149	31.79	2.559	0.016
5	0.339	0.126	38.39	2.685	0.011
6	0.254	0.116	38.75	2.186	0.034
7	0.177	0.127	35.12	1.395	0.172
8	0.310	0.121	33.69	2.560	0.015
9	0.209	0.116	39.29	1.805	0.079
10	0.277	0.115	37.18	2.397	0.022

Is the quality of strategy usage related to learning gains and pupil diameter?

Because the difference in test performance between self-explanation and paraphrasing was associated with greater cognitive processing (as measured by pupil diameter), analyses were performed to determine if the quality of strategy usage was related to learning gains and cognitive processing. Audio recordings were used to determine the number of times each SERT technique was used for each paragraph in the paraphrasing and self-explanation conditions. First, transcriptions were segmented into single statements and then each statement was coded as a paraphrase, prediction, monitoring statement, bridging statement, or elaboration statement. Because research indicates that learning from text is most highly correlated with the bridging and elaboration techniques (McNamara, 2004), self-explanation quality was measured as: a) the proportion of any technique except paraphrasing (i.e., not the lowest level strategy); and b) the proportion of statements that were bridging and elaboration statements (i.e., the highest level strategies). Each of these measures was a proportion of total number of statements per paragraph excluding statements that were rereading of the text. Due to technical failure with the audio recording equipment, one participant was excluded from the analyses.

Because the relationship between learning gains and pupil diameter was significant only in the later intervals (4-10), analyses of the relationship between strategy usage and cognitive processing were limited to these intervals. Pearson correlations between each strategy quality measure and average pupil diameter for each slide during the strategy portion of each paragraph were calculated for each participant; because the correlation is calculated within each participant, differences in overall pupil diameter across participants were accounted for. A one-sample *t*-test was then used to determine if the average correlation across participants was different from zero. For texts in the paraphrase condition, the correlation between the proportion of non-paraphrase self-explanations and pupil diameter ($M = -.11$, $SD = .20$) was

significantly different from 0, $t(20) = -2.39$, $p = .027$. For texts in the self-explanation condition, the correlation between the proportion of non-paraphrase self-explanations and pupil diameter ($M = -.15$, $SD = .36$) was significantly different from 0, $t(20) = -2.25$, $p = .036$. The findings suggest that greater usage of non-paraphrase self-explanations was associated with smaller pupil diameter for both the paraphrase and self-explanation conditions.

The within-participant correlation analyses with pupil diameter were repeated focusing on the proportion of the two most effective individual techniques, bridging and elaboration. Six participants did not use elaboration and were removed from the analyses for that technique. The correlation between bridging usage and pupil diameter ($M = -.23$, $SD = .33$) was significantly different from 0, $t(20) = -3.14$, $p = .005$. The correlation between elaboration usage and pupil diameter ($M = .22$, $SD = .24$) was also significantly different from 0, $t(14) = 3.48$, $p = .004$. In sum, greater usage of elaborations was associated with greater pupil diameters, but greater usage of bridging was associated with smaller pupil diameters.

Finally, the mind wandering ratings were examined to determine if they were related to pupil diameter or strategy quality. The same analysis procedure was followed where correlations for each individual were calculated between mind wandering rating, pupil diameter during reading, pupil diameter during strategy performance, and strategy quality. These correlations were then tested in a one-sample *t*-test to see if they were different from zero. None of the correlations were significantly different from zero likely reflecting that mind wandering was on average relatively low in the task with mean ratings ranging from 2.8-3.2 across the three reading strategies.

Discussion

The purpose of this study was first to determine whether differences in the use of effective reading strategies could be tracked using pupillometry. As the data in Figure 1 show, pupil diameter shows large differences between the least effective strategy, rereading, and the other two strategies. In terms of theory, this difference was predicted based on prior research showing that rereading engages the brain's cognitive control network less than paraphrasing and self-explanation (Moss et al., 2013, 2011). Pragmatically, the association with pupil diameter and strategy is important given the much broader set of circumstances that can make use of pupillometry relative to the other physiological measures showing strategy differences (i.e., fMRI).

Theoretically more interesting was that pupil diameter discriminates between paraphrasing and self-explanation. Neuroimaging studies showed that these two strategies engage cognitive control brain regions to a similar degree. However, self-explanation additionally engages regions in the brain involved in building a coherent representation of the text, including regions involved in inferencing and retrieval of memories (Moss & Schunn, 2015). Therefore, the current results suggest that pupil diameter is tracking engagement of at least some of these non-cognitive control regions in addition to cognitive control regions.

In addition to generally tracking the difference between reading strategies, the current results also indicate that pupil diameter is related to the degree to which self-explanation is effectively utilised as well as how much is learned from the text. If the additional pupil diameter increase for self-explanation over paraphrasing is due to the engagement of inferencing and elaboration processes to build a more coherent representation of the text, then it would be expected that the difference in pupil diameter between self-explanation and paraphrasing would be related to the learning gain difference between these two strategies. This was indeed the case especially after the first few seconds of applying the strategy, as shown in Table 2.

However, an investigation of the two most prominent SERT techniques, bridging and elaboration, and their relationship to pupil diameter shows that these two techniques have different effects on pupil diameter. Elaboration was positively associated with pupil diameter and bridging was negatively correlated with pupil diameter. While self-explanation was generally associated with a higher average pupil diameter, there is interesting variation in how the components of self-explanation are associated with pupil diameter. To better understand these effects, it is important to understand how neural activity is associated with pupil diameter.

Pupil diameter and the locus coeruleus-norepinephrine system

While pupil diameter has in the past been used as a measure of cognitive load (e.g., Beatty, 1982; Peysakhovich et al., 2015; Querino et al., 2015), it is likely that a simple notion of load does not capture the underlying mechanism driving changes in pupil diameter. Pupil diameter is known to be highly correlated with activity in the LC-NE system. The LC-NE is a brainstem neuromodulatory nucleus that sends projections broadly throughout the cortex (Aston-Jones & Cohen, 2005). One theory of LC-NE function is that it modulates neural activity throughout the cortex to promote adaptive behaviour including periods of low tonic activity punctuated by periods of phasic increases of activity upon encountering task-relevant stimuli (Aston-Jones & Cohen, 2005). In other words, this system promotes task engagements via relatively low baseline activity that periodically increases in activity when relevant stimuli are detected by a perceptual processing system in the brain in order to allow the detected stimulus to effect activity in response-related regions.

However, Aston-Jones and Cohen (2005) also identify another mode of the LC-NE system that is associated with higher levels of baseline activity thought to be engaged when the current task has relatively low utility. This higher tonic level of activity is argued to be involved in searching for alternative activities or other approaches to the task to increase the value of the outcome. Another theoretical explanation for this mode is that it involves decoupling from the perceptual processing stream in order to mentalise or think about one's own thoughts (Frith & Frith, 2006; Smallwood et al., 2011). Tonic increases in pupil diameter have been found to be associated with a lack of phasic responses and periods of mind wandering (Smallwood et al., 2011). This mentalising is often associated with contemplating one's own thoughts or those of others (i.e., theory of mind), but brain regions associated with mentalising are also consistently found to activate during discourse comprehension (Ferstl, Neumann, Bogler, & von Cramon, 2008; Moss & Schunn, 2015). It is likely that mentalising engages brain regions associated with retrieving knowledge and forming connections between retrieved knowledge. These processes are also required for discourse comprehension.

Therefore, the increase in pupil diameter associated with self-explanation is likely caused by period phasic increases in pupil diameter and an increase in baseline diameter associated with the higher tonic LC-NE activity associated with more internally directed attention or decoupling from the perceptual environment. Different SERT techniques are also likely to involve different degrees of externally-directed versus internally-directed attention. Elaboration involves the connection of text information with retrieved knowledge and therefore is likely to be more associated with a tonic increase in pupil diameter as memory is searched for relevant knowledge and semantic connections are made to that knowledge. Bridging, by contrast, is concerned with connecting information across sentences in the text or between information contained in diagrams to that contained in the text, and therefore it is more likely to involve externally-directed attention or at least a greater proportion of external-to-internal attention. This line of reasoning may explain why elaboration was associated with increases in pupil diameter, but bridging was associated with decreases in average pupil diameter.

Other effects that can be explained with this understanding of pupil diameter include the pattern of results for rereading in Figure 1. Rereading shows a rapid decrease in pupil diameter starting in the second interval. This decrease in mean pupil diameter likely occurs because of reduced phasic activity reflecting the relatively low levels of cognitive control needed for rereading. Rereading also shows an increase in pupil diameter during the final interval likely reflecting the phasic response caused by the cue presented on the screen to stop rereading.

While other studies have found effects of mind wandering on pupil diameter (Smallwood et al., 2011), they have used experience sampling probes to more closely localise the mind wandering events in time. Therefore the lack of mind wandering results in the current study is likely due to the poor temporal resolution of the mind wandering prompt that participants responded to. The primary focus of the current experiment was on detecting differences between reading strategies, and the mind wandering measure was included to assess whether such a simple prompt would be sufficient for examining some mind wandering effects.

Limitations

The current study provides evidence that pupil dilation could be used to determine reading strategy, but there are some limitations in the use of pupillometry in an everyday educational setting. Though eye trackers with the ability to measure pupil dilation are more affordable and easier to implement than MRI scanners, they can still be financially burdensome for schools. Most schools may not be able to afford multiple eye trackers or the software required to analyse pupil dilation. However, developments in technology may make this technology more affordable, including embedding eye-tracking into laptops.

Another issue that affects the use of eye tracking outside of a laboratory setting is the collection of accurate data. In laboratory settings, accommodations are made for potential environmental factors, such as having fixed, optimal lighting conditions. In a classroom setting, environmental control is more difficult. Further, in the current study, the conditions that participants performed the current study are similar to a tutoring session for an individual child rather than in a large classroom context. Further research is needed to determine how the presence of other participants performing the same task, which is typical of a classroom setting, may influence the ability to determine reading strategy using pupil dilation.

In addition to the technological limitations, the sample used in the current study—college students—may not be the ones that would benefit the most from the current findings. A potential application of the current study is to aid in the improvement of reading strategy usage in elementary and middle-school children. However, eye movement data of younger children tend to have more noise which could make it more difficult to determine what pupil dilation change is in response to (Whiteside, 1974). As a result, additional data filtering may be needed when working with younger children.

Finally, the results of this study should be replicated with a larger sample size such that learning gains between self-explanation and paraphrasing can be clearly shown. The lack of an interaction between reading strategy and pupil diameter may be due to lack of statistical power rather than the lack of a moderating effect of reading strategy on the relationship between pupil diameter and learning gains.

Future directions in applying pupillometry to reading comprehension

The results presented here indicate that pupil diameter could be used to track effective reading strategy use as a more cost-effective means of providing feedback on the reading of difficult expository texts. For example, a computer tutoring system could use pupil diameter to estimate the quality of reading strategy usage without having to have detailed representations of the text being read as is currently the case with other reading strategy tutors. In addition, such a system would not require the learner to produce typewritten self-explanations that would be seen as more effortful and discourage use of the strategy.

However it is also clear that simply tracking average pupil diameter does not capture all of the available and useful dynamics in the pupil diameter data. It may be possible to develop more advanced algorithms that monitor for long-term increases in pupil diameter as well as shorter phasic increases to better track the use of different kinds of SERT techniques. In conclusion, the data presented here show that there are interesting changes in pupil diameter associated with strategy use that can serve as the basis for future studies that examine the relationship between pupil diameter dynamics and the resolution of comprehension difficulties via the application of effective strategies such as self-explanation.

Acknowledgments

References

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*(1), 403–450. http://dx.doi.org/10.1146/annurev.neuro.28.061604.135709Barr, D. J., Levy, R., Scheepers, C., &

Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. http://dx.doi.org/10.18637/jss.v067.i01

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276-292. http://dx.doi.org/10.1037/0033-2909.91.2.276

Callender, A. A., & McDaniel, M. A. (2009). The limited benefits of rereading educational texts. *Contemporary Educational Psychology*, *34*(1), 30-41. http://dx.doi.org/10.1016/j.cedpsych.2008.07.001

Carrier, L. M. (2003). College students' choices of study strategies. *Perceptual and Motor Skills, 96*(1), 54-56. http://dx.doi.org/10.2466/pms.2003.96.1.54

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*(2), 145-182. http://dx.doi.org/10.1207/s15516709cog1302_1

Chi, M. T. H., De Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439-477. http://dx.doi.org/10.1207/s15516709cog1803_3

Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research*, *34*(3), 315-346. http://dx.doi.org/10.1207/S15327906MBR3403_2

Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, *29*(5), 581–593. http://dx.doi.org/10.1002/hbm.20422

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–534. http://dx.doi.org/10.1016/j.neuron.2006.05.001

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective & Behavioral Neuroscience*, *10*(2), 252–269. http://dx.doi.org/10.3758/CABN.10.2.252

Graesser, A. C., McNamara, D. S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iSTART. *Educational Psychologist*, *40*(4), 225-234. http://dx.doi.org/10.1207/s15326985ep4004_4

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122. http://dx.doi.org/10.1037/0033-295X.99.1.122

Karpicke, J. D., Butler, A. C., & Roediger III, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, *17*(4), 471-479. http://dx.doi.org/10.1080/09658210802647009

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmerobjects of lme4 package). R package Version 2.0 –3 [Computer software]. Retrieved from https://CRAN.R-project.org/package=lmerTest

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476-490. http://dx.doi.org/10.3758/BF03210951

Marx, J. D., & Cummings, K. (2007). Normalized change. *American Journal of Physics*, *75*(1), 87-91. http://dx.doi.org/10.1119/1.2372468

McDaniel, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, *14*(2), 200-206. http://dx.doi.org/10.3758/BF03194052

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, *38*(1), 1–30. http://dx.doi.org/10.1207/s15326950dp3801_1

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*(3), 247–288. http://dx.doi.org/10.1080/01638539609544975

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*(1), 1–43. http://dx.doi.org/10.1207/s1532690xci1401_1

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 222-233. http://dx.doi.org/10.3758/BF03195567

McNamara, D. S., O'Reilly, T. P., Best, R. M., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research*, *34*(2), 147-171. http://dx.doi.org/10.2190/1RU5-HDTJ-A5C8-JVWE

Moss, J., & Schunn, C. D. (2015). Comprehension through explanation as the interaction of the brain's coherence and cognitive control networks. *Frontiers in Human Neuroscience*, *9*, 562. http://dx.doi.org/10.3389/fnhum.2015.00562

Moss, J., Schunn, C. D., Schneider, W., & McNamara, D. S. (2013). The nature of mind wandering during reading varies with the cognitive control demands of the reading strategy. *Brain Research*, *1539*, 48–60. http://dx.doi.org/10.1016/j.brainres.2013.09.047

Moss, J., Schunn, C. D., Schneider, W., McNamara, D. S., & VanLehn, K. (2011). The neural correlates of strategic reading comprehension: Cognitive control and discourse comprehension. *NeuroImage*, *58*(2), 675–686. http://dx.doi.org/10.1016/j.neuroimage.2011.06.034

Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'Connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus–noradrenergic arousal function in humans. *Psychophysiology*, *48*(11), 1532–1543. http://dx.doi.org/10.1111/j.1469-8986.2011.01226.x

Peysakhovich, V., Causse, M., Scannella, S., & Dehais, F. (2015). Frequency analysis of a task-evoked pupillary response: Luminance-independent measure of mental effort. *International Journal of Psychophysiology*, *97*(1), 30-37. http://dx.doi.org/10.1016/j.ijpsycho.2015.04.019

Querino, E., dos Santos, L., Ginani, G., Nicolau, E., Miranda, D., Romano-Silva, M., & Malloy-Diniz, L. (2015). Cognitive effort and pupil dilation in controlled and automatic processes. *Translational Neuroscience*, *6*(1), 168-173. http://dx.doi.org/10.1515/tnsci-2015-0017

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, *10*(3), 241–255. http://dx.doi.org/10.1207/s1532799xssr1003_3

Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., … Schooler, J. W. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PLOS ONE*, *6*(3), e18298. http://dx.doi.org/10.1371/journal.pone.0018298

Whiteside, J. A. (1974). Eye movements of children, adults, and elderly persons during inspection of dot patterns. *Journal of Experimental Child Psychology, 18*(2), 313–332. http://dx.doi.org/10.1016/0022-0965(74)90111-8

Corresponding author: Aaron Wong, ayw12@msstate.edu