

9-4-2015

Evaluating Pupil Dilation as a Measure of Working Memory and Logical Thinking Manuscript

Ryan Zahran
Grand Valley State University

Nathan Barrows
Grand Valley State University

Jessica VandenPlas
Grand Valley State University

Follow this and additional works at: <http://scholarworks.gvsu.edu/sss>

 Part of the [Medicine and Health Sciences Commons](#)

Recommended Citation

Zahran, Ryan; Barrows, Nathan; and VandenPlas, Jessica, "Evaluating Pupil Dilation as a Measure of Working Memory and Logical Thinking Manuscript" (2015). *Student Summer Scholars*. 155.
<http://scholarworks.gvsu.edu/sss/155>

This Open Access is brought to you for free and open access by the Undergraduate Research and Creative Practice at ScholarWorks@GVSU. It has been accepted for inclusion in Student Summer Scholars by an authorized administrator of ScholarWorks@GVSU. For more information, please contact scholarworks@gvsu.edu.

Evaluating Pupil Dilation as a Measure of Working Memory and Logical Thinking Manuscript

Introduction

College-level general chemistry courses often have the unfortunate reputation of requiring extensive memorization and recall of disparate and disconnected trivia. However, the chemical education research literature suggests that students in these courses struggle for a variety of reasons unrelated to recall including poor preparation in previous chemistry courses, deficient algebraic skills, and working memory limitations. In educational research, “working memory” is a psychological construct that indicates a person’s ability to simultaneously hold and manipulate information. Although various theories of working memory abound, they share a common prediction with respect to problem solving: if the working memory demand of a question is greater than the problem solver’s working memory capacity, he or she will fail to answer the question correctly. Consequently, questions with high working memory demands tend to assess students’ working memory capacities rather than their content knowledge (Niaz, 1996) (Tsapralis, Kousathana, & Niaz, 1998). As students’ grades should reflect their content knowledge and not a cognitive trait, it is essential for instructors to monitor the working memory demands of their assessment items.

Unfortunately, the historical methods of measuring working memory demand and capacity are unsatisfactory. First, assessments of questions’ working memory demands typically involve performing a task analysis (e.g., counting the discrete steps in solving a problem) from an expert’s perspective. Values determined in this manner assume that the student takes the same problem-solving route of the expert, which is highly unlikely given that experts and novices categorize and approach questions differently (Chi, Feltovich, & Glaser, 1981). Second, working memory capacity is often measured via digit-span recall tests (e.g., recounting a string of digits in reverse order) given immediately before or soon after a problem-solving session. Such tests cannot be given while a student is problem solving as they are disruptive. Recent research has explored the utility of noninvasive monitoring of physiological responses as measures of working memory.

Previous research has drawn correlations between pupil diameter and ones load on memory. It has been discovered that there is a direct correlation between the size of ones pupil and their cognitive load (Kahneman, & Beatty, 1966). They found that the greater the number of digits each participant had to recall, the greater the diameter of their pupil was. This is referred to as the task-evoked pupillary response (TEPR). In light of this discovery, eye-tracking research has begun to use pupil diameter as a measure of working memory. The purpose of this study is to validate the method of using pupil diameter as a direct measurement of working memory demand. This study will use existing measurements of working memory (digit-span recall tests) to correlate to working memory demand. This will also allow future work to use pupil diameter as a real-time measure of working memory demand within a variety of tasks.

Method

Participants

A total of 20 participants took part in the experiment after giving written informed consent, following institutional review board (IRB) regulation of Grand Valley State University. The participants included 6 males and 14 females between the ages of 18 and 22. 1 female participant was excluded from the analysis due to poor data sampling.

Apparatus and stimuli

Forward, and backward, digit span testing, along with digit ordering testing was performed on all participants. Testing was performed in a quiet testing room using a standard PC, controlled by Tobii software. Pupil diameter was recorded by a Tobii T60 eye tracking system (Figure 1), which is built into a 17-inch computer monitor, on which the stimuli were displayed. Cameras and illuminators are hidden behind sunlight blocking filters in the monitor. The Tobii eye tracker hardware uses FDA approved near infrared diodes to produce non-invasive reflection patterns on the users corneas. A camera then collects these reflection patterns along with other user characteristics.

The PC was equipped with two monitors, one visible to the participant and one visible to the experimenter. First, participants sat down and received an explanation of the procedure of the experiment; they were then given the IRB consent form to sign. The participant then filled out a demographic form that assessed the age, gender, and whether they wore contact lenses. This made it possible to evaluate the influences of demographic variables.

The participant was then seated at the Tobii T60 instrument for calibration. The calibration screen asks participants to follow a red dot around the screen while the instrument finds their eyes in order to correlate participant eye position with on-screen fixation location. Participants were shown sample slides of what they were going to be looking at during the testing and verbally explained the procedure. Participants were given directions and examples of each digit test before they were given and given the opportunity to begin the test at their own pace. The digits were prerecorded and given off of a recording device. A “beep” sound was given off of the recording device to signal to the participant both the beginning and end of the digit test. Participants verbally responded to the test approximately 1 second after the digits were read off and an ending “beep” sound was made. During the test participants were asked to fixate their eyes on a small plus symbol in the middle of the screen to measure pupil dilations.

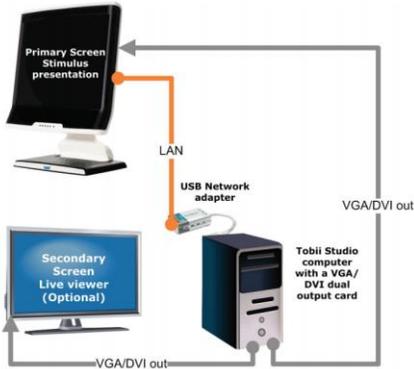


Figure 1. Tobii T60 eye tracker and tracking set-up.

Digit span scoring

The same format for scoring was given to all three of the digit span tests. The participants started with a low amount of digits being given (2 for the DSB, 3 for the DFS and DOT, as shown in Figure 2). They were given two trials for each test and the amount of digits tested were increased by 1 if they verbalized at least one of the trials correctly. The test was ended when the participants failed to complete either of the two trials correctly, or reached the maximum amount of digits (9 for the DSF, 8 for the DSB, and 11 for the DOT). The tests were scored in real time by the investigator and responses were also checked for accuracy by reviewing the audio recording

Code Number: _____

<p>In the following test, you will be presented several digits in random order. Please recall these digits in the same order they are given. <i>Example: 3-2-8; Solution: 3-2-8</i></p> <table border="0"> <tr><td>3</td><td>5 8 2</td><td>6 9 4</td></tr> <tr><td>4</td><td>6 4 3 9</td><td>7 2 8 6</td></tr> <tr><td>5</td><td>4 2 7 3 1</td><td>7 5 8 3 6</td></tr> <tr><td>6</td><td>6 1 9 4 7 3</td><td>3 9 2 4 8 7</td></tr> <tr><td>7</td><td>5 9 1 7 4 2 8</td><td>4 1 7 9 3 8 6</td></tr> <tr><td>8</td><td>5 8 1 9 2 6 4 7</td><td>3 8 2 9 5 1 7 4</td></tr> <tr><td>9</td><td>2 7 5 8 6 2 5 8 4</td><td>7 1 3 9 4 2 5 6 8</td></tr> </table> <p>In the following test, you will be presented several digits in random order. Please recall these digits in the opposite order they are given. <i>Example: 3-2-8; Solution: 8-2-3</i></p> <table border="0"> <tr><td>2</td><td>2 4</td><td>5 8</td></tr> <tr><td>3</td><td>6 2 9</td><td>4 1 5</td></tr> <tr><td>4</td><td>3 2 7 9</td><td>4 9 6 8</td></tr> <tr><td>5</td><td>1 5 2 8 6</td><td>6 1 8 4 3</td></tr> <tr><td>6</td><td>5 3 9 4 1 8</td><td>7 2 4 8 5 6</td></tr> <tr><td>7</td><td>8 1 2 9 3 6 5</td><td>4 7 3 9 1 2 8</td></tr> <tr><td>8</td><td>9 7 3 7 6 2 5 8</td><td>7 2 8 1 9 6 5 3</td></tr> </table>	3	5 8 2	6 9 4	4	6 4 3 9	7 2 8 6	5	4 2 7 3 1	7 5 8 3 6	6	6 1 9 4 7 3	3 9 2 4 8 7	7	5 9 1 7 4 2 8	4 1 7 9 3 8 6	8	5 8 1 9 2 6 4 7	3 8 2 9 5 1 7 4	9	2 7 5 8 6 2 5 8 4	7 1 3 9 4 2 5 6 8	2	2 4	5 8	3	6 2 9	4 1 5	4	3 2 7 9	4 9 6 8	5	1 5 2 8 6	6 1 8 4 3	6	5 3 9 4 1 8	7 2 4 8 5 6	7	8 1 2 9 3 6 5	4 7 3 9 1 2 8	8	9 7 3 7 6 2 5 8	7 2 8 1 9 6 5 3	<p>In the following test, you will be presented several digits in random order. Please recall these digits in ascending order immediately after presentation. If digits occur twice, recall them twice as well. <i>Example: 5-2-8-2; Solution: 2-2-5-8</i></p> <table border="0"> <tr><td>3</td><td>3 7 2 (2 3 7)</td><td>6 1 6 (1 6 6)</td></tr> <tr><td>4</td><td>8 4 7 3 (3 4 7 8)</td><td>7 2 7 6 (2 6 7 7)</td></tr> <tr><td>5</td><td>2 6 5 3 1 (1 2 3 5 6)</td><td>5 1 8 5 2 (1 2 5 5 8)</td></tr> <tr><td>6</td><td>4 8 3 9 2 7 (2 3 4 7 8 9)</td><td>3 0 3 7 2 5 (0 2 3 3 5 7)</td></tr> <tr><td>7</td><td>7 4 3 1 8 2 6 (1 2 3 4 6 7 8)</td><td>9 6 5 2 8 6 3 (2 3 5 6 6 8 9)</td></tr> <tr><td>8</td><td>3 7 6 1 8 2 5 0 (0 1 2 3 5 6 7 8)</td><td>2 6 3 2 1 8 7 5 (1 2 2 3 5 6 7 8)</td></tr> <tr><td>9</td><td>7 4 2 9 0 1 3 8 6 0 1 2 3 4 6 7 8 9</td><td>8 2 0 3 1 7 4 2 9 0 1 2 2 3 4 7 8 9</td></tr> <tr><td>10</td><td>9 5 4 8 1 3 0 5 8 6 0 1 3 4 5 5 6 8 8 9</td><td>7 4 2 9 8 4 0 3 7 5 0 2 3 4 4 5 7 7 8 9</td></tr> <tr><td>11</td><td>3 7 5 2 1 8 3 0 4 7 9 (0 1 2 3 3 4 5 7 7 8 9)</td><td>8 5 4 0 3 7 9 4 2 6 5 (0 2 3 4 4 5 5 6 7 8 9)</td></tr> </table>	3	3 7 2 (2 3 7)	6 1 6 (1 6 6)	4	8 4 7 3 (3 4 7 8)	7 2 7 6 (2 6 7 7)	5	2 6 5 3 1 (1 2 3 5 6)	5 1 8 5 2 (1 2 5 5 8)	6	4 8 3 9 2 7 (2 3 4 7 8 9)	3 0 3 7 2 5 (0 2 3 3 5 7)	7	7 4 3 1 8 2 6 (1 2 3 4 6 7 8)	9 6 5 2 8 6 3 (2 3 5 6 6 8 9)	8	3 7 6 1 8 2 5 0 (0 1 2 3 5 6 7 8)	2 6 3 2 1 8 7 5 (1 2 2 3 5 6 7 8)	9	7 4 2 9 0 1 3 8 6 0 1 2 3 4 6 7 8 9	8 2 0 3 1 7 4 2 9 0 1 2 2 3 4 7 8 9	10	9 5 4 8 1 3 0 5 8 6 0 1 3 4 5 5 6 8 8 9	7 4 2 9 8 4 0 3 7 5 0 2 3 4 4 5 7 7 8 9	11	3 7 5 2 1 8 3 0 4 7 9 (0 1 2 3 3 4 5 7 7 8 9)	8 5 4 0 3 7 9 4 2 6 5 (0 2 3 4 4 5 5 6 7 8 9)
3	5 8 2	6 9 4																																																																				
4	6 4 3 9	7 2 8 6																																																																				
5	4 2 7 3 1	7 5 8 3 6																																																																				
6	6 1 9 4 7 3	3 9 2 4 8 7																																																																				
7	5 9 1 7 4 2 8	4 1 7 9 3 8 6																																																																				
8	5 8 1 9 2 6 4 7	3 8 2 9 5 1 7 4																																																																				
9	2 7 5 8 6 2 5 8 4	7 1 3 9 4 2 5 6 8																																																																				
2	2 4	5 8																																																																				
3	6 2 9	4 1 5																																																																				
4	3 2 7 9	4 9 6 8																																																																				
5	1 5 2 8 6	6 1 8 4 3																																																																				
6	5 3 9 4 1 8	7 2 4 8 5 6																																																																				
7	8 1 2 9 3 6 5	4 7 3 9 1 2 8																																																																				
8	9 7 3 7 6 2 5 8	7 2 8 1 9 6 5 3																																																																				
3	3 7 2 (2 3 7)	6 1 6 (1 6 6)																																																																				
4	8 4 7 3 (3 4 7 8)	7 2 7 6 (2 6 7 7)																																																																				
5	2 6 5 3 1 (1 2 3 5 6)	5 1 8 5 2 (1 2 5 5 8)																																																																				
6	4 8 3 9 2 7 (2 3 4 7 8 9)	3 0 3 7 2 5 (0 2 3 3 5 7)																																																																				
7	7 4 3 1 8 2 6 (1 2 3 4 6 7 8)	9 6 5 2 8 6 3 (2 3 5 6 6 8 9)																																																																				
8	3 7 6 1 8 2 5 0 (0 1 2 3 5 6 7 8)	2 6 3 2 1 8 7 5 (1 2 2 3 5 6 7 8)																																																																				
9	7 4 2 9 0 1 3 8 6 0 1 2 3 4 6 7 8 9	8 2 0 3 1 7 4 2 9 0 1 2 2 3 4 7 8 9																																																																				
10	9 5 4 8 1 3 0 5 8 6 0 1 3 4 5 5 6 8 8 9	7 4 2 9 8 4 0 3 7 5 0 2 3 4 4 5 7 7 8 9																																																																				
11	3 7 5 2 1 8 3 0 4 7 9 (0 1 2 3 3 4 5 7 7 8 9)	8 5 4 0 3 7 9 4 2 6 5 (0 2 3 4 4 5 5 6 7 8 9)																																																																				

Figure 2. Digit span scoring sheet

Note: The directions before each test were presented on the screen to each participant before the test. Each column in the sheet represents the two trials that were given to each participant.

Data analysis

All failed or uncompleted trials were removed from the data set. The number of participants correctly completing each trial are given in Table 1.

The Tobii software measures the validity of each eye measurement from a 0-4 scale (“0 if the eye is found and the tracking quality is good” 4 being the worst measurement (User Manual Revision 4)); all measurements with a validity 1-4 were removed from the data set due to poor measurement. All data sets (pupil diameter vs time) were graphed, and outliers were removed via visual inspection (12 points removed).

Table 1. Number of Participants Successfully Completing Each Digit Test

Digits	DSF	DSB	DOTA
2	*	20	*
3	20	20	20
4	20	18	19
5	20	13	17
6	18	5	14
7	13	3	6
8	6	1	5
9	2	*	4
10	*	*	2
11	*	*	1
12	*	*	0

*Digits not included in this test

The diameter was then transformed into percent dilation. Because the actual size of the pupil varies between individuals, this allowed for a normalization of pupil diameter data. Percent dilation was calculated compared to a baseline pupil diameter measurement. Baseline was determined by taking the average pupil diameter while participants were fixated at a neutral screen for 10 seconds before all tests. The diameter for the first 5 seconds was removed to reduce poor data caused by a novelty effect. The formula for this transformation is shown below:

$$\frac{x - y}{y} * 100 = \text{percent change in pupil diameter}$$

x = max diameter (mm) for a given trial

y = average diameter of baseline

For each trial, the maximum percentage dilation achieved by a participant was then recorded. Maximum pupil diameter was chosen as the variable of interest because it should correlate with the highest load experienced during a trial by an individual.

A three-way analysis of variance (ANOVA) test was conducted to compare the maximum percent dilation of each individual to the type of digit span test, the number of digits given, and the trial of the test (1st or 2nd).

Results

The ANOVA test revealed that no significant interactions effects were found. The main effects that were investigated were test effect, digit effect, and practice (trial) effect.

Test Effect

The ANOVA analysis revealed that significant differences exist among the three digit span tests ($F_{2,444}=3.56$, $p=0.029$). There was a significant difference between the DSF and DOT test, but the DSB was not significantly different than the other two tests. The maximum pupil dilation reached during each test (averaged for all participants over all trials), is given in Table 2.

Table 2. Participant Average of Maximum Pupil Dilation by Test

Test	N	Mean	Std. Deviation
DOT*	158	20.57	10.87
DSB	144	19.44	11.17
DSF*	184	17.94	10.15

*Statistically significantly different according to Fisher's least significant difference (LSD) test.

Digit Effect

The average maximum pupil dilation achieved during recall of digit spans of varying lengths is shown for each test in Figures 3-5. Pupil dilation was found to be significantly different based on number of digits given in a trial ($F_{8,444}=1.96, p=0.050$). A digit effect is observed where the percent change in pupil diameter increases as a higher digit is given. A post-hoc test shows a significant difference lies between digit 10 and digits 3,4,5.

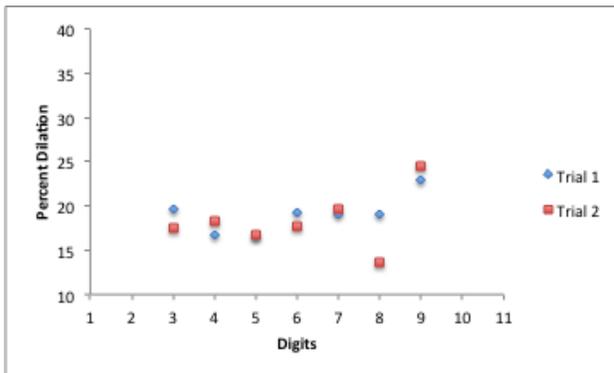


Figure 3. Digit Span Forward Test: Participant Average of Maximum Pupil Dilation by Digit Span

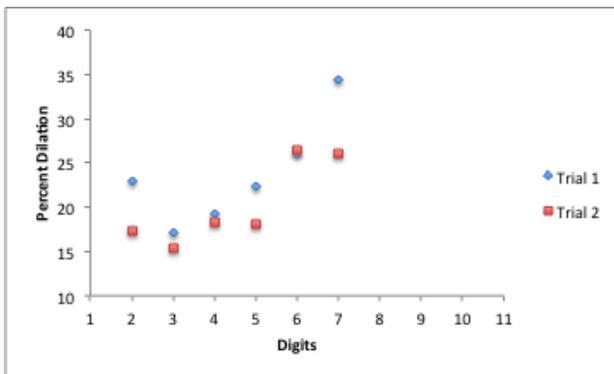


Figure 4. Digit Span Backward Test: Participant Average of Maximum Pupil Dilation by Digit Span

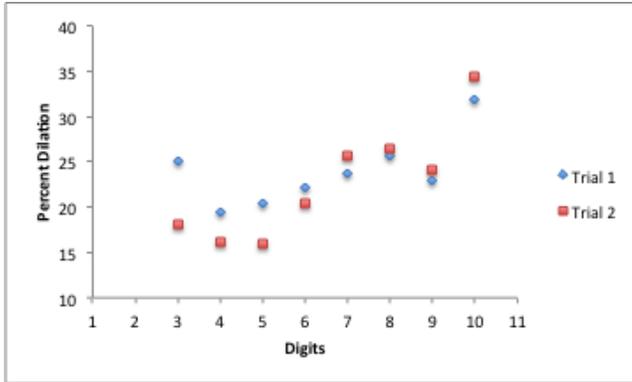


Figure 5. Digit Ordering Test: Participant Average of Maximum Pupil Dilatation by Digit Span

Practice Effect

When averaged over all tests and all digit spans, differences in pupil diameter can be seen between the first and second trial (Table 3); however this difference is not statistically significant.

Table 3. Participant Average of Maximum Pupil Dilatation by Trial

Trials	N	Mean	Std. Deviation
1	244	20.35	10.91
2	242	18.13	10.46

Qualitative Observations

For a given test, participants completed 2 trials of each digit length. Each time the digit length was increased by one, the first digit of this trial appeared to be unusually high, compared to other digits within this trial, and compared to the first digit of the subsequent trial of matching digit length. (Figures 3, 4, and 5). This observation may be due to a novelty effect, in which the first trial of a new test may be significantly more difficult than subsequent trials.

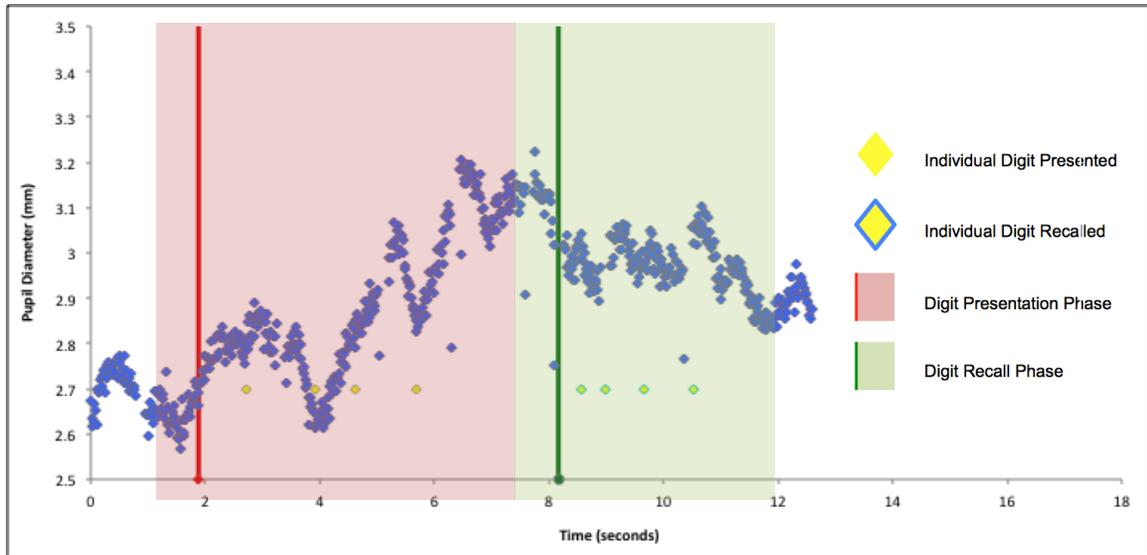


Figure 6. Pupil diameter during Digit Ordering Test of Participant 10.

Peaks were observed in pupil diameter immediately following presentation or recall of each digit during all three digit span tests (Figure 6). The number of peaks during recall appears to vary according to reported participant strategy. Participants who reported “chunking” numbers (grouping numbers in memory) showed the same number of peaks during the listening phase as the number of “chunks” they remembered. However, during the listening phase, the number of peaks did not always match the number of chunks the participant reported. This has yet to be analyzed but it seems to be due to participants changing the method in which they remember the digits.

Conclusions

The digit-ordering test shows a greater TEPR than the digit forward test. This suggests that the digit-ordering test requires a greater amount of working memory than the digit forward test. This may be because the DOTA required transformation of the digits, while the DSF required simple recall.

For all tests, the greater number of digits given in a trial also showed a higher TEPR. This was expected because a greater number of digits would require a greater working memory load and thus a higher TEPR, which is consistent with the hypothesis. The biggest differences were between digits 3,4,5 and 10. All of the digits were not significantly significant from each other perhaps due to small sample size. If larger sample sizes were taken, it is possible that all of the digits would show bigger differences from each other.

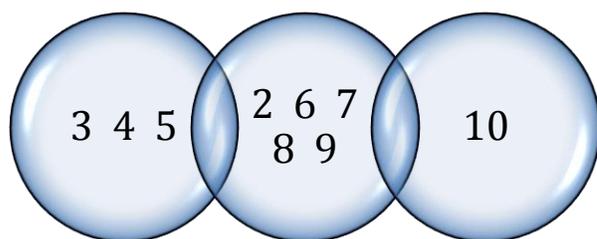


Figure 6. Venn diagram of statistically significant digits during all three digit span tests. In addition, pupil diameter appears to decrease on subsequent trials of each digit span. Although these differences are not statically significant, this may be due to small sample size, large standard deviations, or limited number of trials. A significant difference between trials would not be surprising because of a novelty effect. The subsequent trials may be easier due to practice from the first trial.

These results suggest that percentage dilation is a valid measure of working memory demand. In cases where higher demand was expected—more difficult tests, increased digit spans, novel trials—increased pupil dilation was also seen.

Future Work

Results of this project can go many directions. Future research includes identifying a smoothing algorithm to remove noise and to further clean data and possibly reduce standard deviations, data collection using a larger sample size with greater number of trials to better interpret results, and the identification of other variables that may serve as a better marker of TERP than percentage dilation eg. integrate area under curve, identify individual peak maxima, etc. Other variables may reveal why some of the results were inconsistent, such as why all digits weren't different than one another.

This research has a variety of practical uses such in educational research. Methods used in this project can be used on tests of problem solving to try to identify working memory demand in real-time during different phases of problem solving and during the solution of different types of problems. This research also has the potential to be used in marketing research. Pupil diameter can be used to measure how much information people can take in through a commercial, website, etc. Pupil diameter is a good study of working memory and currently has many educational applications and it needs to be further studied for future applications.

References

- Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science* , 5, 121-152.
- Kahneman, D., & Beatty, J. (1966). Pupil Diameter and Load on Memory. *Science*, 1583-1585.
- Niaz, M. (1996). Reasoning strategies of students in solving chemistry problems as a function of developmental level, functional M capacity and disembedding ability. *International Journal of Science Education* , 18 (5), 525-541.
- User manual--Tobii Studio Cersion 3.2. Tobii Technology, 2012
- Tsaparlis, G., Kousathana, M., & Niaz, M. (1998). Molecular-equilibrium problems: Manipulation of logical structure and of M-demand, and their effect on student performance. *Science Education* , 82, 437-454.