

Exploring Differences Among Student Populations During Climate Graph Reading Tasks: An Eye Tracking Study

Rachel M. Atkins, North Carolina State University, USA
Karen S. McNeal, Auburn University, USA

ABSTRACT

Communicating climate information is challenging due to the interdisciplinary nature of the topic along with compounding cognitive and affective learning challenges. Graphs are a common representation used by scientists to communicate evidence of climate change. However, it is important to identify how and why individuals on the continuum of expertise navigate graphical data differently as this has implications for effective communication of this information. We collected and analyzed eye-tracking metrics of geoscience graduate students and novice undergraduate students while viewing graphs displaying climate information. Our findings indicate that during fact-extraction tasks, novice undergraduates focus proportionally more attention on the question, title and axes graph elements, whereas geoscience graduate students spend proportionally more time viewing and interpreting data. This same finding was enhanced during extrapolation tasks. Undergraduate novices were also more likely to describe general trends, while graduate students identified more specific patterns. Undergraduates who performed high on the pre-test measuring graphing skill, viewed graphs more similar to graduate students than their peers who performed lower on the pre-test.

Keywords: Geoscience Education Research; Geocognition; Eye-Tracking; Graph Reading

The average temperature of our planet has increased 1.5 degrees F over the past 100 years and is projected to increase another 0.5 to 8.6 degrees F over the next century (U.S. Environmental Protection Agency, 2014). This projected increase in temperature is alarming due to the potential harmful impacts on water supplies, agriculture, power and transportation systems, the natural environment, and human health and safety (IPCC (United Nations), 2013; U.S. Environmental Protection Agency, 2014; U.S. Global Change Research Program (USGCRP), 2016). Over 97% of scientists publishing about climate topics agree that current global warming is anthropogenic (human-induced) (Cook et al., 2007). Despite overwhelming scientific agreement about the increase in temperature and evidence indicating anthropogenic effects on our climate, there is less agreement about anthropogenic climate change among non-scientist Americans than among non-scientists in most other countries (Weber & Stern, 2011). Worldwide, level of education is the most accurate predictor of climate change awareness (Lee, Markowitz, Howe, Ko, & Leiserowitz, 2015). Additionally, the public's concern (Leiserowitz, 2005) and risk perceptions (Kahan et al., 2012) vary with level of knowledge about climate science (Leiserowitz, Smith, & Marlon, 2010). However, depending on cultural worldview, increased knowledge may or may not increase risk perceptions of climate change (Aksit, McNeal, Gold, Libarkin, & Harris, 2017; Kahan, Jenkins-Smith, & Braman, 2011). For example, people tend to align their climate risk perceptions with their peers who share the same societal views (Kahan et al., 2011). Moreover, political and religious affiliations often impact an individual's knowledge and perceptions about climate change (McNeal, Walker, & Rutherford, 2014). Many students hold misconceptions about the underlying processes of Earth's climate system (Lombardi & Sinatra, 2012; McNeal, Spry, Mitra, & Tipton, 2014; Niebert & Gropengiesser, 2012; Shepardson, 2011; Sterman & Sweeney, 2002). For example, students will often mistake singular weather events as climatic representations (Lombardi & Sinatra, 2012). The dynamics of the greenhouse effect are also commonly misunderstood along with the impacts that the greenhouse effect has on climate (Gautier, Deutsch, & Rebich, 2006; Libarkin, Thomas, & Ording, 2015; Rebich & Gautier, 2005; Shepardson, 2011). Deep time, as it relates to the past and future, is also among the topics that students either fail to acknowledge or misunderstand (Libarkin, Kurdziel, & Anderson, 2007; Lombardi & Sinatra, 2012; Rebich & Gautier, 2005). This deep time understanding is of particular importance when understanding the differences between weather and climate along with interpreting

graphs depicting rates of change over long periods of time. These knowledge deficits among students indicate a need for increased emphasis on and explanation of climate education.

While researchers work to understand our climate and improve future climate predictions, the communication of this research is typically not tailored to non-scientists (Somerville & Hassol, 2011). The challenges of climate communication have been investigated to help scientists tailor climate-related messages to appropriate audiences (Bostrom, Bohm, & O'Connor, 2013) and research has shown that the use of imagery can be an effective way to promote understanding of climate change along with influencing feelings about preventing negative impacts that can result (O'Neill, Boykoff, Niemeyer, & Day, 2013). Graphs have been found to be the most effective way of communicating the climate consensus of researchers (van der Linden, Leiserowitz, Feinberg, & Maibach, 2014). Concepts related to the essential principles of climate literacy such as changes in sea level, temperature and greenhouse gas concentrations over long periods of time (National Oceanic and Atmospheric Administration (NOAA), 2009) are frequently communicated using graphs (Winn, 1987). While some perceive a graphical representation of information to be easier to interpret, it can also trigger complex cognitive processing if not designed appropriately (Carpenter & Shah, 1998; Huang, Eades, & Hong, 2009; Korner, 2011). In order to create more effective graphs and scaffold student interpretations during classroom learning, it is important to understand how they attend to these graphs.

The interpretation of scientific graphs has been deemed as a necessary skill for achieving scientific literacy among students, however, little research in graph interpretation in science (as opposed to math) has been conducted (Glazer, 2011). Research focused on graph interpretation is especially limited at the secondary and tertiary education levels (Glazer, 2011) as well as in the content area of the geosciences. Recent work in geoscience education has indicated that the design of the Intergovernmental Panel on Climate Change (IPCC) generated climate change graphs needs to be re-considered in order to better cognitively scaffold the viewer's ability to interpret the complex information contained in these visuals (Harold, Lorenzoni, Shipley, & Coventry, 2016). Additionally, high school level classroom implementations have shown that direct instruction guiding students' interpretation of climate change graphs have led to improvements in student performance (Rule & Meyer, 2009). However, research is quite limited in the geosciences regarding how viewer attention is distributed while interpreting climate change graphs, and what the differences are, if any, between novices and experts when viewing these graphs.

Researchers aiming to understand graph comprehension have developed conceptual models that explain the processes involved in graph comprehension. These models incorporate top down (knowledge-driven) and bottom up (stimulus-driven) processes. Top down elements include the viewer's content familiarity and graph comprehension skills while bottom up features tend to include visual features (e.g., color and line thickness) or graph format (e.g., line versus bar graphs) (Fabrikant, Hespanha, & Hegarty, 2010; Freedman & Shah, 2002; Kriz & Hegarty, 2007; Shah & Freedman, 2011; Shah, Freedman, & Vekiri, 2006; Trickett & Trafton, 2006). Within this framework there are three types of processes viewers engage in when making inferences from graphical displays: (i) identifying graph features and making judgements about them, such as determining the slope of a line (Carpenter & Shah, 1998; Freedman & Shah, 2002; Okan, Galesic, & Garcia-Retamero, 2016), (ii) translating the visual features into conceptual relations, which can include determining differences in spatial features such as variations between bar heights (Carpenter & Shah, 1998; Freedman & Shah, 2002; Okan et al., 2016), and (iii) identifying specific variables shown in the graph and their numerical values, such as inferring information from the title, axes, labels and legend of a graph (Carpenter & Shah, 1998; Freedman & Shah, 2002; Okan et al., 2016). In addition to these processes, there are individual differences that also play a role in graph literacy. An individual's overall ability to construct or interpret a graph is influenced by their graph skills (i.e. their familiarity with x and y axes locations, familiarity with type of graph displayed, etc.). Those with more education and experience with graphs exhibit higher graphical literacy and are able to store and access information in their long-term memory, allowing them to access knowledge about the properties of graphs more readily, resulting in more accurate interpretation (Okan et al., 2016; Shah & Freedman, 2011). Greater differences in graph comprehension have been observed between more and less experienced individuals, particularly when more complex graphs are used, with graph complexity evoking data extrapolation and inferences as opposed to simple fact-extraction (Maltese, Harsh, & Svetina, 2015; Shah & Freedman, 2011). However, more graph comprehension research has been completed with the use of simple graphs (Ratwani & Trafton, 2008; Shah et al., 2006; Trafton et al., 2000) than with complex graphs (Maltese et al., 2015; Shah & Freedman, 2011). As such, the range task type (e.g., main idea and trend description, fact extraction, extrapolation, etc.) asked of participants during viewing of complex graphs, has also been limited in previous research.

Eye movements are often an involuntary response that can be measured and analyzed to determine engagement with a visual stimulus. Eye tracking has been used as a method for investigating questions in a variety of domains including reading comprehension (Rayner, 1998, 2009), pupillary response and gaze path patterns (Mitra, McNeal, & Bondell, 2017; Resnick, Kastens, & Shipley, 2018), differences in attentional processes using multimedia stimuli (van Gog & Scheiter, 2010), hierarchical map navigation (Korner, 2004) and decision-making (Muldner, Christopherson, Atkinson, & Bursleson, 2009). Eye tracking has also been applied to graph reading and comprehension studies. An expert-novice study revealed that the viewing patterns of college science students differed from non-science students when comprehending questions and looking at various areas of science-related graphs. This study found that science students spent more time reading the question than non-science students; they suggest that science students spent more time “interpreting the questions thoroughly” (Yen, Lee, & Yang, 2012). Okan et al. (2016) used eye-tracking to understand the individual differences in graph comprehension with graphs containing health-related information. Their findings indicated that the viewers’ level of graph literacy predicted performance on graph comprehension and view times of essential information contained within the displays, where higher graph literacy correlated positively with time spent on critical elements and more accurate interpretations.

Eye tracking has been applied to various climate-related topics. Ho, Tsai, Wang and Tsai (2014) explored how prior knowledge affects viewing of climate text and graphics. Students with higher levels of prior knowledge focused longer periods of attention on areas of interest than students with low prior knowledge. Beattie & McGuire (2012) investigated view patterns of individuals while viewing iconic images of climate change based on their attitudes towards the environment. They found that individuals had strong positive, implicit attitudes towards low carbon footprint products were more likely to focus their attention on negative images of environmental damage and climate change than on positive images. McNeal et al. (2014) used eye tracking to understand how students navigated and used the EarthLabs materials, an on-line climate change curriculum for high school students, finding that students engaged mainly with text, not the images on the web-pages and had a particularly difficult time engaging with graphs depicting change over time. However, they also found that most participants found charts, graphs and questions *embedded* in text to be most useful and that attention to text decreased over time, suggesting that a reduction of text on web-based curriculum might increase overall engagement.

This study employs an exploratory research design to understand how introductory undergraduate Earth Systems Science students (undergraduates) allocate their attention while viewing climate graphs. Undergraduate eye movements were compared to a more expert-like population of geoscience graduate students (graduates) who had more experience viewing these types of graphs and completing similar tasks, confirmed by significantly higher scores graphical pre-test. Questions we investigated in this study included:

- 1) How do graduate and undergraduate eye movements differ during various climate graph related tasks?
- 2) To what extent does the amount of time spent viewing various features of climate graphs differ between graduate and undergraduate students?
- 3) In what ways do graduate and undergraduate verbal interpretations of climate change graphs differ?

METHODS

Population

The 58 participants in this study consisted of 45 undergraduates and 13 graduate students. On the continuum of expertise, undergraduates trend toward the novice end of the spectrum and are distinguished from graduate students by their experiences with reading and interpreting graphs, confirmed by their lower scores on the graphical pre-test ($p=0.001$, grad=81.22%, undergrad=69.81%). The undergraduate population consisted of introductory students enrolled in an Earth Systems Science course. Once obtaining human subject’s research approval from the Institutional Review Board (IRB), undergraduate students were recruited from an introductory Earth Systems Science course at a large land-grant university in the southeastern United States. A pre-test was administered at the beginning of the course and participants completed the eye-tracking part of the study by the 4th week of the semester, prior to climate content being covered in the course. Our undergraduate participants were 57.8% male and 42.2% female with a median age of 19 and ranged in class years from freshman (42.2%), sophomore (26.7%), junior (11.1%), senior (13.3%) and fifth-year senior (6.7%). Ethnicities included white/Caucasian (86.7%), black/African/African American (6.7%), Asian/Asian American (2.2%) and other (4.4%). Experts were all 23 (23.1%) or older (76.9%), masters (38.5%) or

PhD (61.5%) students, 38.5% male, 61.5% female and 100% white/Caucasian. Graduate students were either geology or atmospheric science students from the same institution with at least a year of graduate experience prior to participating in this study. The researcher gave brief recruitment presentations to two lectures of undergraduates enrolled in different sections of introductory Earth Systems Science taught by different instructors asking students to participate in an eye-tracking study outside of class. Undergraduates received \$20 in Amazon gift cards as compensation after completion of the pre-test and a 20-30 minute eye-tracking study.

Experimental Design

This study used four different graphs varying in type and complexity (Figure 1). Graph names and abbreviations are summarized in Table 1. Six questions were asked of varying difficulty for each graph (Table 1) (Curcio, 1987; Friel, Curcio, & Bright, 2001). Questions 1 and 2 were open-ended questions asking participants to provide a summary of information displayed or for the identification of a data trend. Questions 3a and 3b required fact extraction from the graph and questions 4a and 4b required participants to make inferences or extrapolate data, which may have involved the use of domain-specific knowledge. All questions involved different tasks, potentially impacting eye-patterns, and were therefore analyzed separately. Participants were asked to view graphs and answer questions using the concurrent verbal protocol (Bojko, 2013). All graphs contain climate content and were slightly modified from their original Environmental Protection Agency (U.S. Environmental Protection Agency, 2014) published version for consistency in color, line thickness and overall readability. The topics of the four graphs covered global greenhouse gas concentrations, sea level/global temperature change, temperature anomalies in the US, worldwide temperature anomalies/carbon dioxide concentrations.

Participants were given an overview of the study before beginning. They were all aware that eye tracking was taking place and that their body movements needed to be limited in order to prevent disruption of data collection. The researcher explained that they would be viewing graphs and answering questions. Each graph was shown for a 5-second free viewing period without a task. This was done to address the need for viewers to orientate themselves with the provided information so that during the question asking period information could be processed (Korner, 2011). Once questions appeared on the screen, participants were asked to think out loud as they answered questions. There was no time limit to each question task, which was accounted for during data analysis by normalizing for time differences. The participant advanced at their own pace to allow for complete thought development and graph exploration. The researcher was present in the room in order to answer any questions during the study. Participants were not informed of research questions or that we were investigating graduate/undergraduate differences prior to data collection.

Figure 1. a) Global Greenhouse Gas Concentration graph (GGG), b) Sea Level and Global Temperature Change graph (SLGT), c) Temperature Anomaly in the Contiguous 48 States graph (TA48), d) Worldwide Temperature Anomaly and Carbon Dioxide Concentration graph (WWTA)

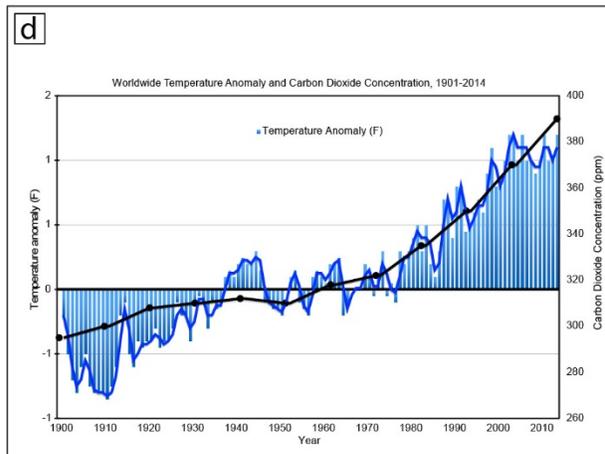
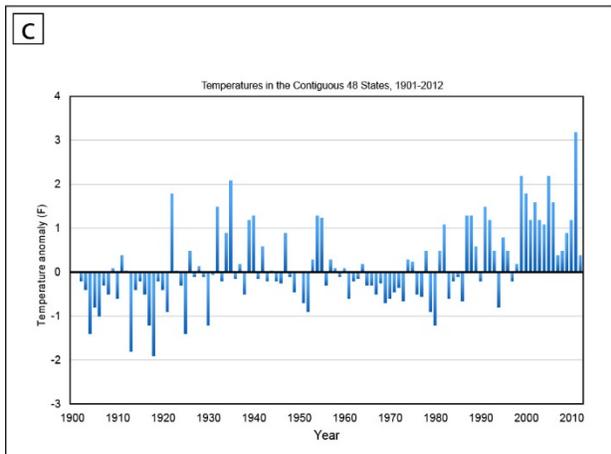
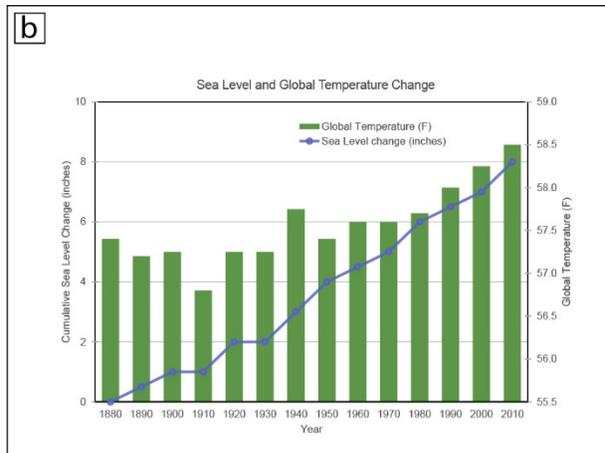
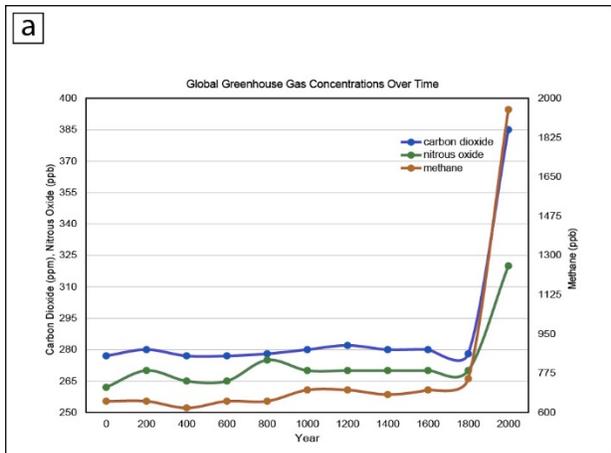


Table 1. Summary of graphs and questions

Abbreviation	Question
Title: Global greenhouse gas concentrations over time (GGG)	
Type: Line graph, 3 variables, 2 y-axes with different scales	
Q1	What is the main idea of this graph?
Q2	What trends do you see in this graph?
Q3a	In 2000, what was the approximate CO ₂ concentration?
Q3b	Over the last two millennium, what was the approximate change in CO ₂ ?
Q4a	What type of activities took place between 1800 and 2000 to cause the sharp rise in global greenhouse gases?
Q4b	What could be a possible explanation for nitrous oxide not increasing as rapidly as the other greenhouse gases?
Title: Sea level and global temperature change (SLGT)	
Type: Line + bar graph, 2 y-axes with different scales	
Q1	What is the main idea of this graph?
Q2	What trends do you see in this graph?
Q3a	In 1950, what was the approximate sea level change?
Q3b	When do global temperature and sea level appear to have the closest correlation?
Q4a	What might global sea level change be in the year 2050?
Q4b	What might global temperature be in the year 2050?
Title: Temperature anomaly in the contiguous 48 states, 1901-2012 (TA48)	
Type: Bar graph, single y-axis	
Q1	What is the main idea of this graph?
Q2	What trends do you see in this graph?
Q3a	How many years has the temperature anomaly been more than 2 degrees F above average?
Q3b	During the first half of the century, were there more or less positive temperature anomalies as compared to the second half?
Q4a	What do you think the temperature anomaly graphs for other countries would look like?
Q4b	Based on the graph, should we expect to see more positive or negative temperature anomalies in the future?
Title: Worldwide temperature anomaly and CO ₂ concentration, 1901-2014 (WWTA)	
Type: Line + bar graph, 2 y-axes with different scales	
Q1	What is the main idea of this graph?
Q2	What trends do you see in this graph?
Q3a	In 1920, what was the approximate carbon dioxide concentration?
Q3b	Describe the relationship between CO ₂ concentration and temperature anomaly during the first half of the century compared to the second half.
Q4a	What could you infer from this graph about how the temperature will change from average conditions into the future?
Q4b	If this trend continues, what are some impacts that humans might see in the future?

Instrumentation

Eye tracking data was collected using a Tobii TX300 eye tracker fixed to a 23-inch computer monitor, collecting at 300 Hz. Calibration was completed for each participant to ensure accuracy, precision, and consistency within and among participant trials and allows for corrective lenses to be worn without affecting results. Participants were notified about the potential risks of eye tracking and sat ~65cm from the monitor and gazed at the computer screen to view the provided graphs with an unobstructed view. The unobtrusive nature of eye tracking methods allowed for the researcher to capture natural eye movements.

All participants completed a graphing pre-test to determine prior knowledge about graph reading, interpretation and construction (see Supplementary Materials), which was scored out of 34 possible points. Undergraduate scores ranged from 32.35% to 89.29%, with an average score of 69.81% and were additionally classified as low (<67%, n=15), moderate (67% ≤ x < 75%, n=16) or high performers (≥75%, n=14). Graduate student scores ranged from 64.71% to

91.18%, with an average score of 81.22% (n=13). A Mann-Whitney U test indicated a significant difference between undergraduate and graduate student pre-test scores ($p=0.001$).

To address validity issues with the eye tracking study design and the graphical skills pre-test, both were reviewed and revised for content validity by two additional researchers, a climate science expert and a geoscience education researcher. Additionally, two pilot studies were conducted with 16 geoscience experts at a scientific conference and 10 undergraduate REU participants, yielding modifications based on preliminary results of participant performance, ability to understand tasks, and test item comprehension.

DATA ANALYSIS

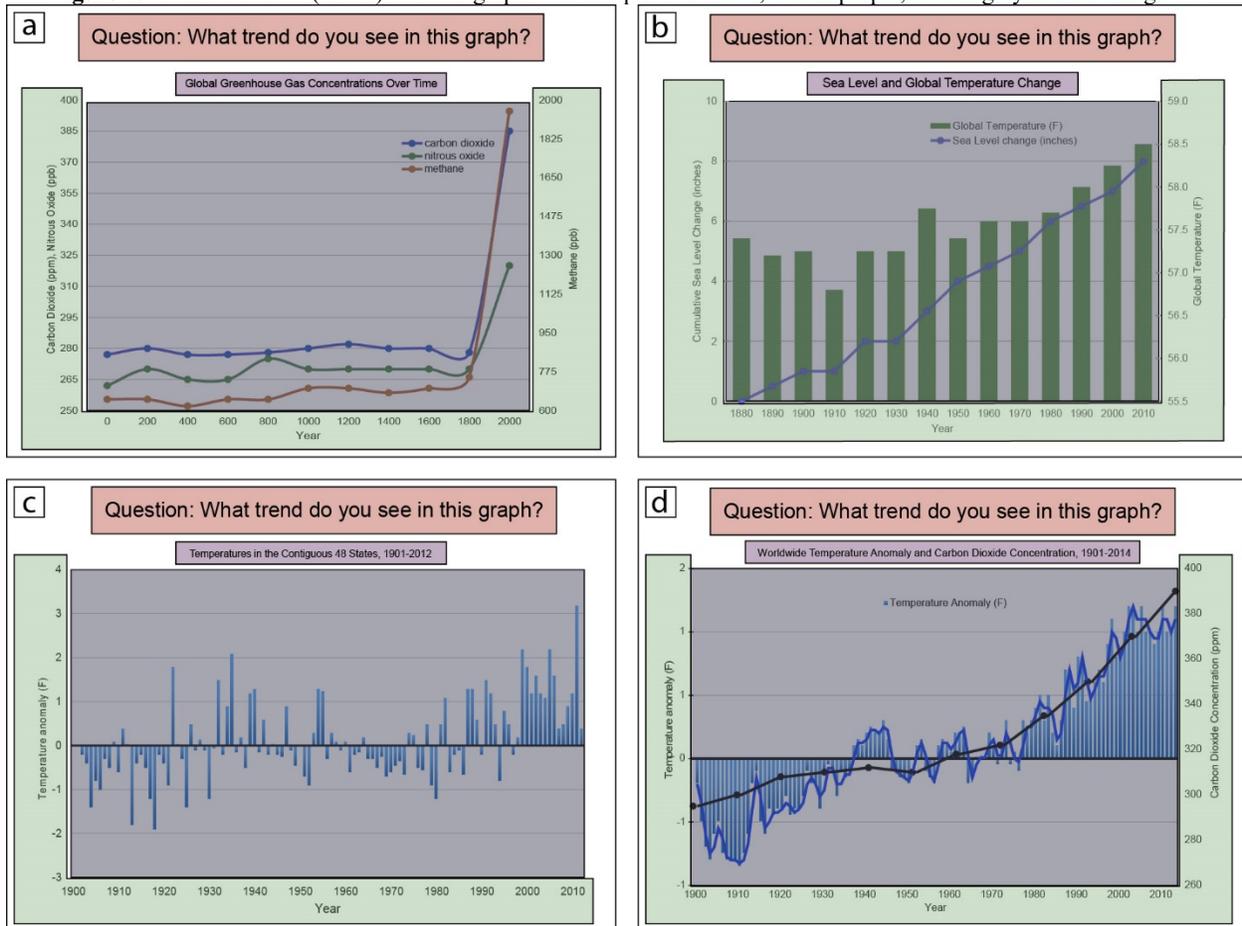
Eye tracking data were analyzed using a variety of quantitative and qualitative approaches. Two aspects of eye movements that are most often studied include saccades and fixations. Saccades are the short periods of rapid eye movement between fixations that redirect participant gaze from one fixation to another (Ramat, Leigh, Zee, Shaikh, & Optican, 2008). These can occur up to four times in a second and participants are effectively blind while they occur (Land, 2012). Fixations are the points between saccades where the eye is nearly stationary for a relatively long period of time (~70-100ms). These eye movements are of particular interest as it is during these points during viewing that processing takes place (Bojko, 2013). Based on our conceptual model of expertise, we used fixation duration and fixation count, *a priori*, as our metrics of interest. Differences in fixations among science majors and non-majors were observed by Yen et al. (2012), therefore, we expected to see a similar pattern with our populations of graduates and undergraduates.

Variables and Areas of Interest

Variables collected for each participant included gender, major (STEM, non-STEM), pre-test performance, race, ethnicity, age and year in school. Gender, age and year in school did not significantly impact view patterns. In addition, major (STEM, non-STEM), race and ethnicity did not have enough diversity within the sample for statistical analysis, therefore, variables used for analysis include experience (graduate or undergraduate) and pre-test performance. Experience was defined by participants' level of education; graduate students had at least one year of experience in their graduate career and novice undergraduate students were freshmen through seniors in college. Novice undergraduates were further categorized and analyzed by their pre-test performance.

Four areas of interest (AOI's) were identified for each graph based on salient features found within these areas. The title, axes, question and data were used to systematically analyze fixations (Figure 2).

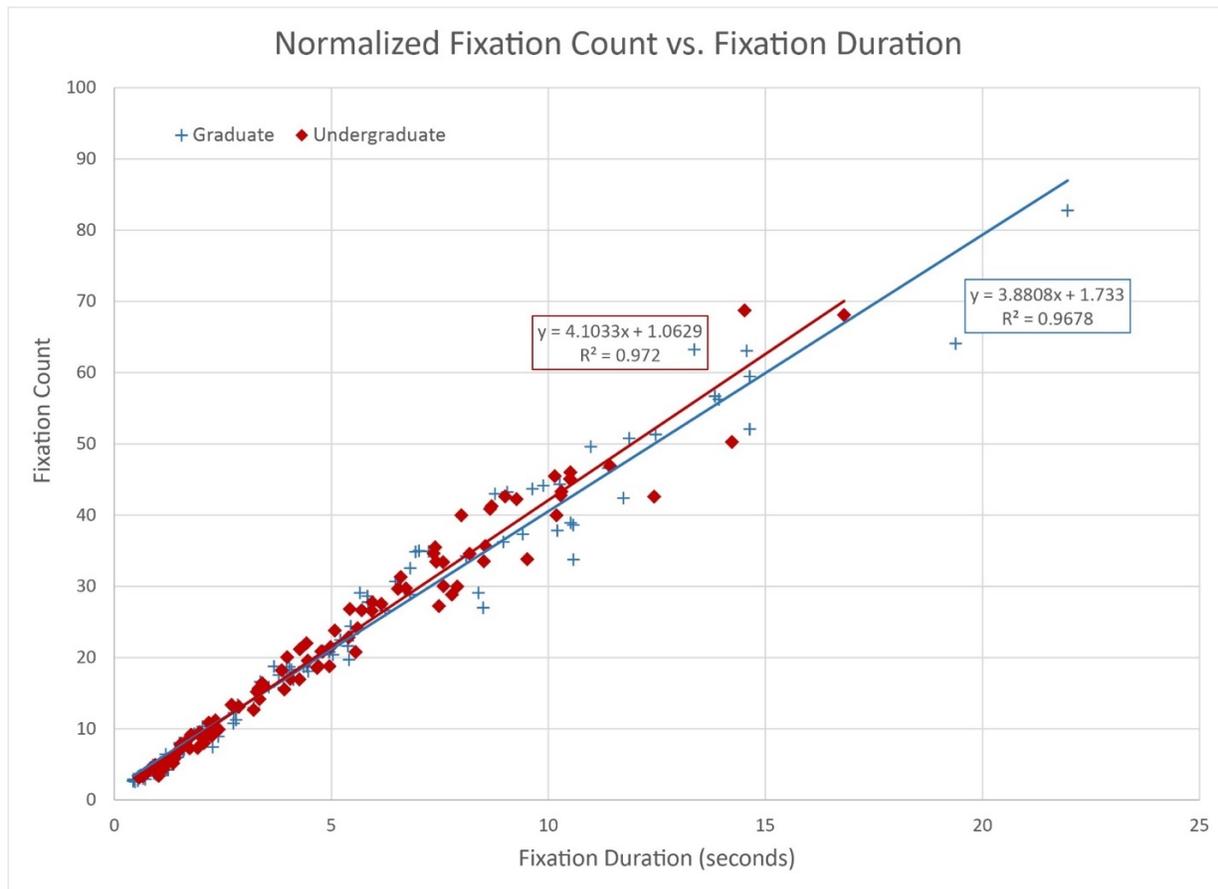
Figure 2. Areas of interest (AOI's) for each graph with the question in red, title in purple, data in gray and axes in green.



QUANTITATIVE ANALYSIS

Quantitative data for this study was analyzed using Tobii Studio 2012, Microsoft Excel 2013, and SPSS version 23.0. Quantitative results from the eye-tracking data were exported from Tobii Studio based on metrics chosen by the researcher. For example, if the interest is in the time it takes for a participant to first fixate on an answer, the researcher must first define the specific area of interest (AOI) in the Tobii software for the answer. Once AOI's have been identified, they are selected and fixation data are exported to excel based on the metric of interest. Fixations can be defined using dispersion-based or velocity-based algorithms, however, for the purpose of this study we used the Tobii I-VT (Velocity-Threshold Identification) fixation filter. This uses the average of the left and right eyes and identifies a fixation when eyes move slower than 30 degrees/second over a period of at least 20ms. Eye movements with duration times of less than 60ms were filtered out as non-fixations as they resemble saccades. Missing data were linearly interpolated for up to 75ms gaps. This 75ms value is shorter than an average blink.

Figure 3. Normalized fixation count vs. fixation duration scatter plot. Each point on the graph represents an average for the indicated population (graduate or undergraduate) for each area of interest for each graph.



Total fixation duration and number of fixations within an area of interest were the metrics selected for analysis prior to data collection, as it is during fixations that cognitive processing occurs. A difference between fixation count and duration could indicate that certain graph features required more cognitive processing than others and this difference would need to be investigated further. For our study, fixation count and duration were very strongly correlated ($R^2 = 0.97$ for undergraduates, $R^2 = 0.97$ for graduates) (Figure 3). Given this strong correlation, we take the approach of Kastens, Shipley, & Boone (2016) and will reference only the fixation duration metric that we call ‘time spent’ for the remainder of this paper.

To account for time differences while participants were viewing graphs, data were normalized for the total amount of time a participant spent on an image. For example, if a participant spent one minute viewing a graph, and 15 of those 60 seconds were spent looking at an AOI, the value for that participant's view time within the AOI would be normalized to 0.25. This allows for comparison among participants who spent differing amounts of time viewing graphs. Data were also normalized for the size of the AOI's, which is needed when making comparisons among multiple AOI's as participants would statistically spend more time in larger AOI's. For this calculation, time normalized values were divided by the pixel area of each AOI. This normalization allows for the comparison between multiple AOI's of various sizes.

In order to investigate relationships between populations, scatterplots were constructed and used to determine differences between graduate and undergraduate view times (Figures 4-7). These were analyzed by graph type, AOI

and question type. Correlation coefficients and Pearson correlations (where applicable) were calculated to determine relationships between data.

Kolmogorov-Smirnov tests were performed to determine normality of data. For graduate/undergraduate comparison, independent sample t-tests were used for normally distributed metrics and Mann-Whitney U-tests were used when metric data was not normal. For a comparison between other variables in the study, ANOVA tests for normally distributed metrics and Kruskal-Wallis tests were used to determine significance for data that were not normally distributed for these variables. In addition to statistical tests, effect size (Cohen's *d*) was also calculated to determine practical significance of results. While statistical significance is a measure of the likeliness of results occurring by chance, effect size determines the magnitude of the differences found and can therefore provide a more practical interpretation of results. For this study, we use 0.2, 0.5 and 0.8 as our small, moderate and large effect sizes respectively (Daly & Cohen, 1978).

Participant Verbal Responses

In addition to eye movements, qualitative responses from think aloud tasks were collected and analyzed. A concurrent verbal protocol approach combined with participant eye-tracking was followed in a similar way to that of (van Gog, Paas, & Van Merriënboer, 2005). The analysis of participant transcripts was completed using a quantitative description approach with NVivo, a software program used for identifying trends in qualitative data. Recordings ranged in length from 5.18 to 25.45 minutes ($M=10.16$, $SD=4.02$, $n=45$) for undergraduates and 7.85 to 15.30 minutes ($M=11.15$, $SD=2.57$, $n=13$) for graduate students. Transcripts were uploaded and organized within NVivo using manual coding to group responses into nodes. Common words used by participants were identified by the program and once categorized by the researcher, were used to create a text search query. Results from this query were used to make inferences and explain quantitative differences between populations. In addition to the common words identified from NVivo, researchers also used direct quotes from participants to understand contextual differences between graduate and undergraduate responses.

Heat Maps

The eye tracker records an x and y coordinate for each fixation and its location can be graphed and overlain on the image being viewed. These images, called heat maps, are used frequently in eye tracking research as a way to visualize participant attention. While heat maps can be helpful for visualization of data and quick analysis, one must be cautious when using them for analysis for many reasons. The first problem arises when using a heat map for comparison of two unequal groups. In order to accurately compare two groups, an equal number of individuals must be used for each visual. Heat maps used for this study controlled for this effect by selecting a representative sample of undergraduates using the median 13 to compare to the 13 graduate students based on their pre-test results. Researchers must also use caution when comparing results of untimed tasks. When an unlimited amount of time is given to view an image, a participant who takes longer to complete the task will have recorded more fixations than a participant who viewed for less time. This was controlled for in our study by normalizing for the amount of time each participant spent on each graph as well as normalizing for the sizes of the areas of interest. Heat maps are included as a supportive visual to assist the reader in understanding the overall data trends observed. As such, they should be viewed as an aid and readers should refer to our more robust combinations of quantitative statistical analyses and verbal response content analysis for complete understanding.

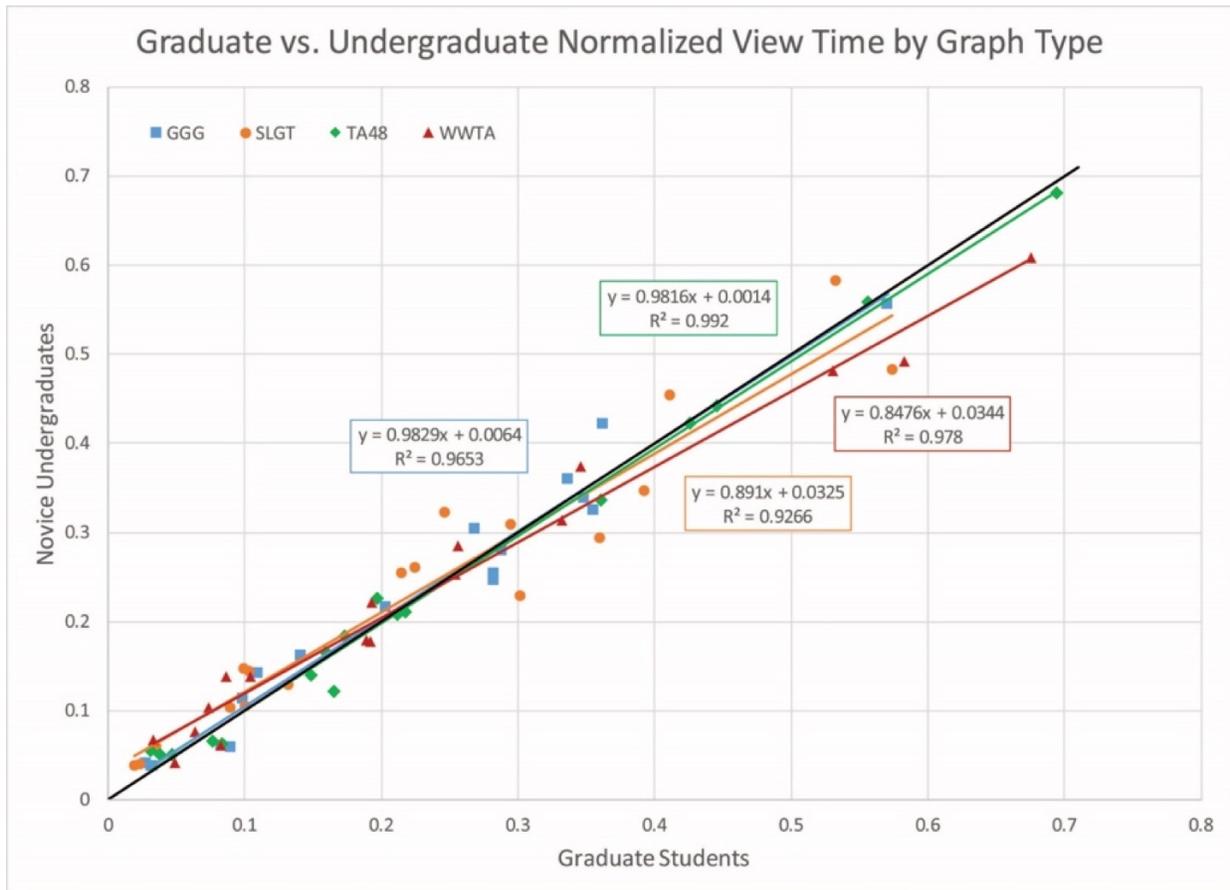
RESULTS

Quantitative Statistical Results

This study asked six questions of four different graphs using two different populations (graduate and undergraduate) and pre-test performance to assess effect on visual patterns while viewing and interpreting graphs. Data were first analyzed by graph type using values normalized for total time spent viewing graphs and the size of each AOI (Figure 4). The x-position of each point represents the proportion of graduate students and the y-position of each point represents the proportion of undergraduate students. Data that plot along the 1:1 line (in black) indicate no difference between the two populations for proportion of attention allocation in an area of interest on each graph. The data cluster along the 1:1 line, indicating little to no consistent differences between graduate and undergraduate view times of the

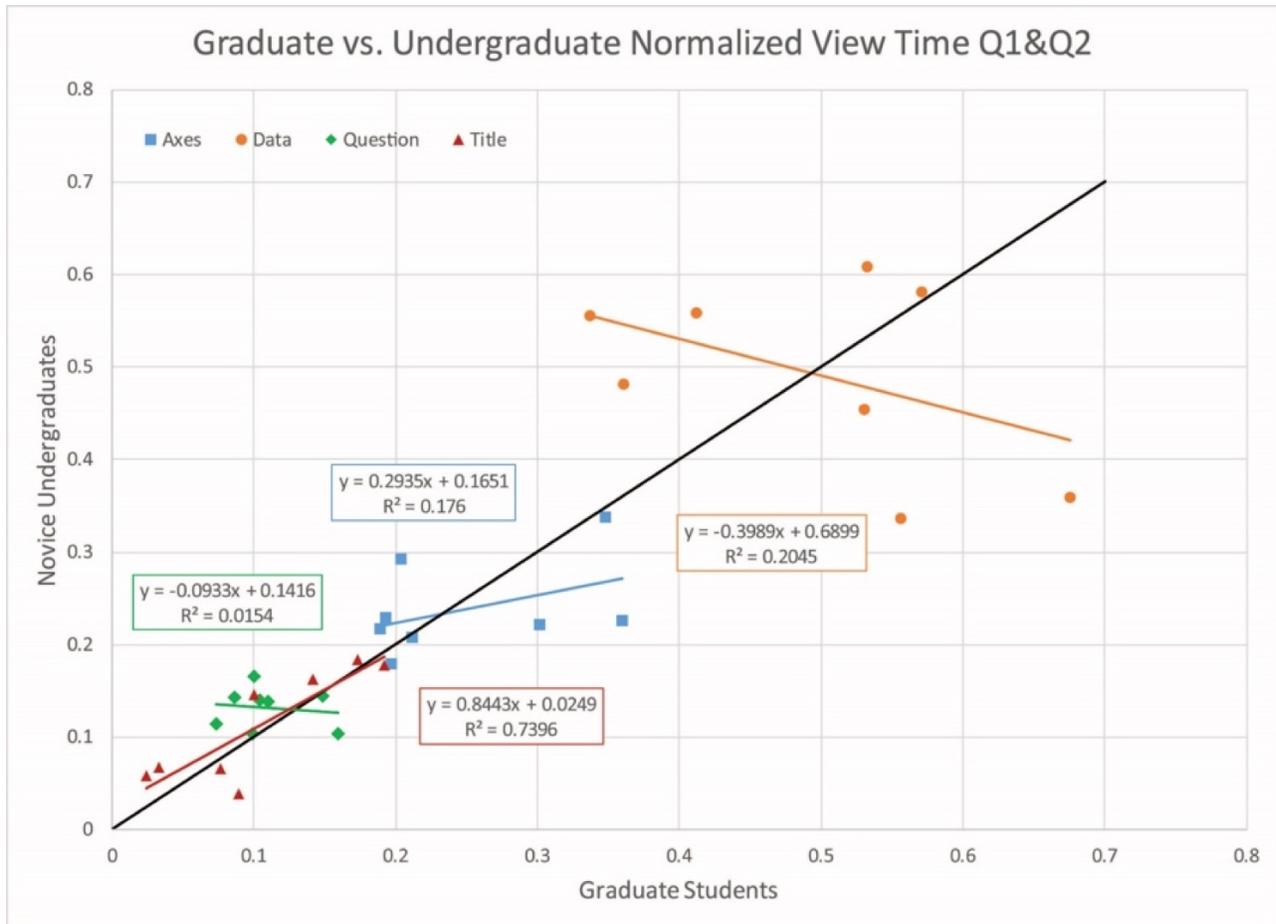
graphs. Given the absence of differences between populations based on the type of graph shown, we compared features (title, axes, data, and question) among graphs.

Figure 4. Graduate vs. Undergraduate normalized view time by graph. Each point represents the average normalized fraction of time spent per pixel (normalization procedure in methods) of graduate and undergraduate students for an area of interest, separated by graph type.



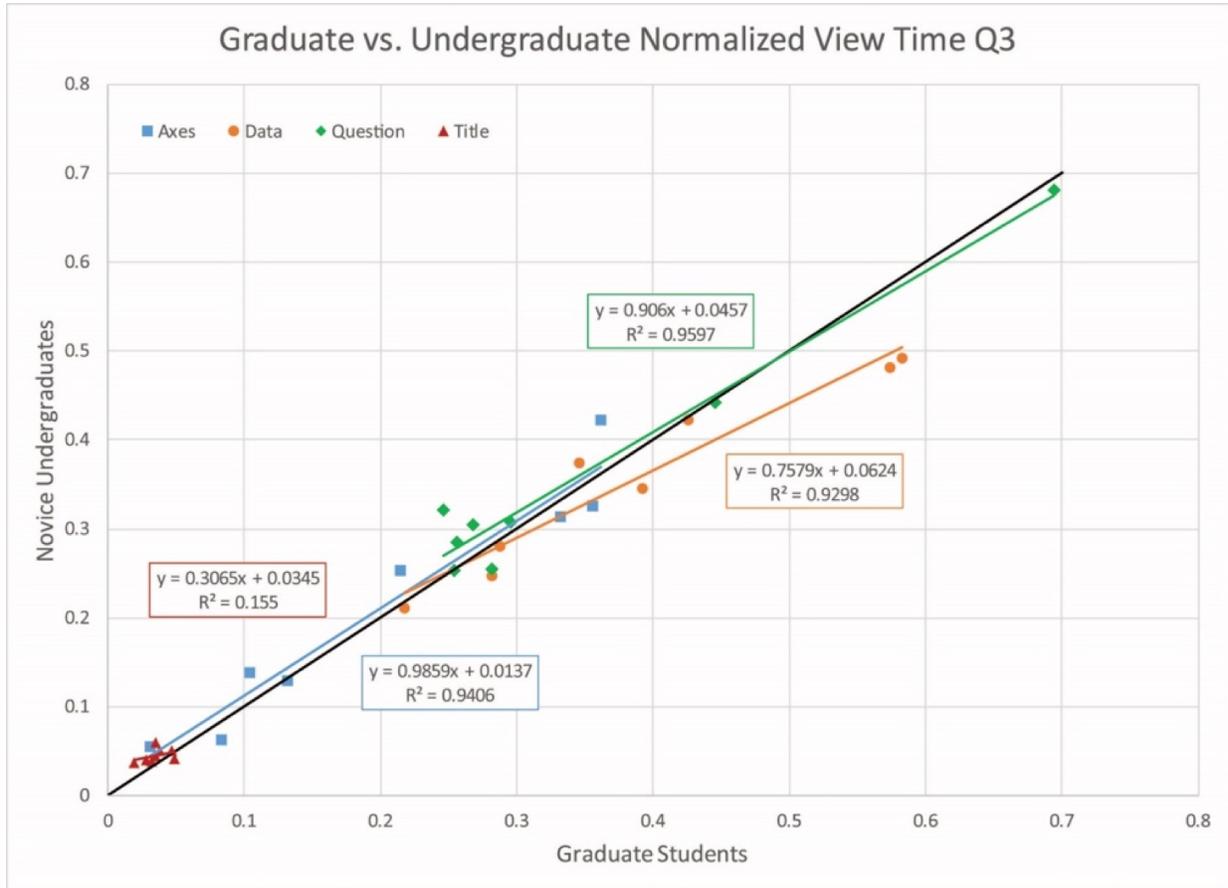
Data were next analyzed based on the type of question being asked (main idea, trend description, fact extraction, extrapolation). Questions 1 and 2 demonstrated variability among graduates and undergraduates with few obvious trends. No statistical significance is found ($p=0.13$), however, practical significance is suggested by a small to moderate effect size (Cohen’s $d=0.38$) for differences in viewing the question, with undergraduates spending more of their time on the question as observed by the higher concentration of question data points located above the 1:1 line (Figure 5).

Figure 5. Graduate vs. Undergraduate normalized view time Q1&2. Each point represents the average normalized fraction of time spent per pixel (normalization procedure in methods) of graduate and undergraduate students for an area of interest, separated by area of interest for questions 1 and 2 on all graphs.



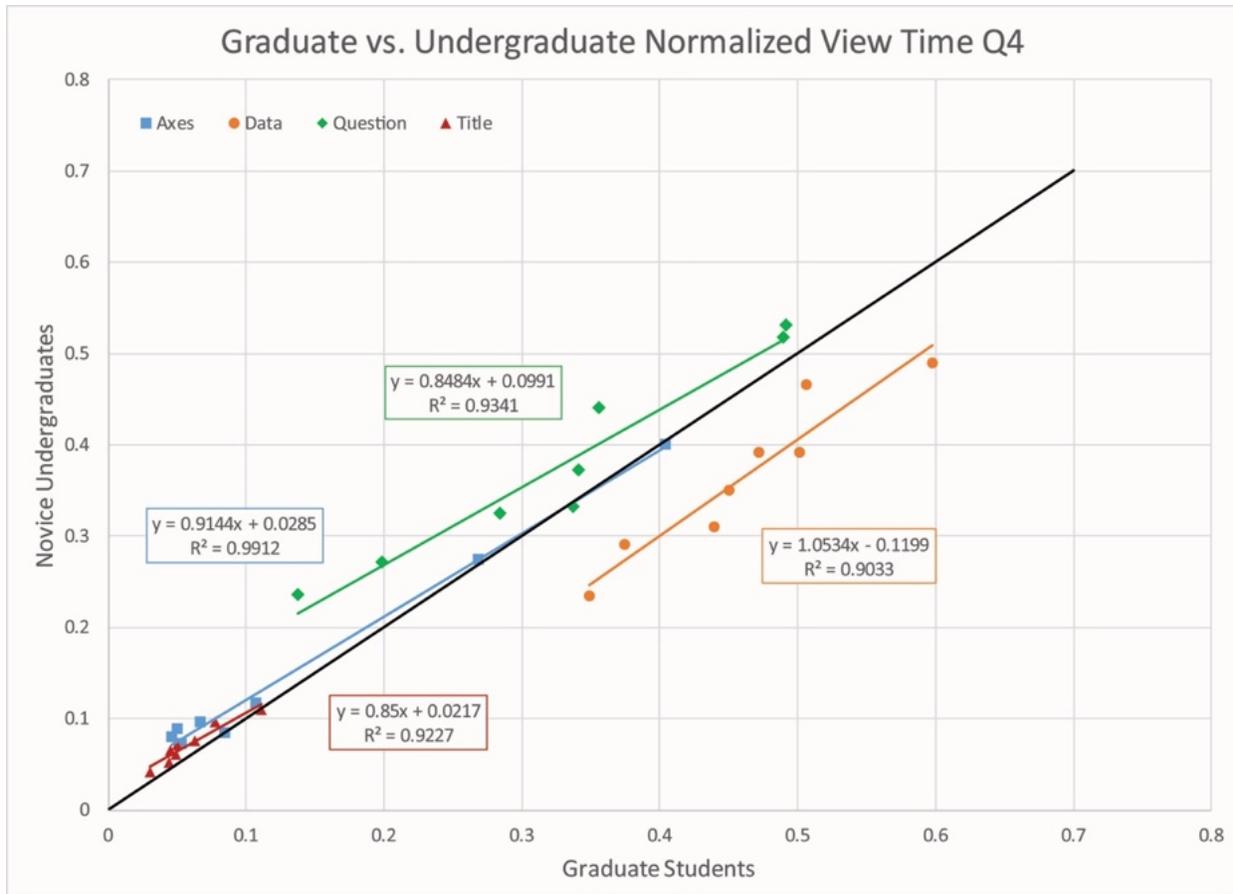
Differences between populations were also observed while viewing question 3 (Figure 6). Practical significance was found with differences in multiple graph elements, indicated by small to moderate effect sizes. These differences are found between graduates and undergraduates when viewing the title ($p=0.06$, Cohen’s $d=0.51$), question ($p=0.35$, Cohen’s $d=0.23$) and axes ($p=0.34$, Cohen’s $d=0.27$). Undergraduates spend more time on all of these elements as shown by their proportional view times in Figure 8.

Figure 6. Graduate vs. Undergraduate normalized view time Q3. Each point represents the average normalized fraction of time spent per pixel (normalization procedure in methods) of graduate and undergraduate students for an area of interest, separated by area of interest for question 3 on all graphs.



The most notable differences are observed during the 4th question task (Figure 7). Undergraduates tend to view the title ($p=0.05$, Cohen’s $d=0.54$), axes ($p=0.24$, Cohen’s $d=0.35$) and question ($p=0.35$, Cohen’s $d=0.25$) longer than graduates (Figure 9). Conversely, graduates spend proportionally more time viewing data ($p=0.21$, Cohen’s $d=0.40$).

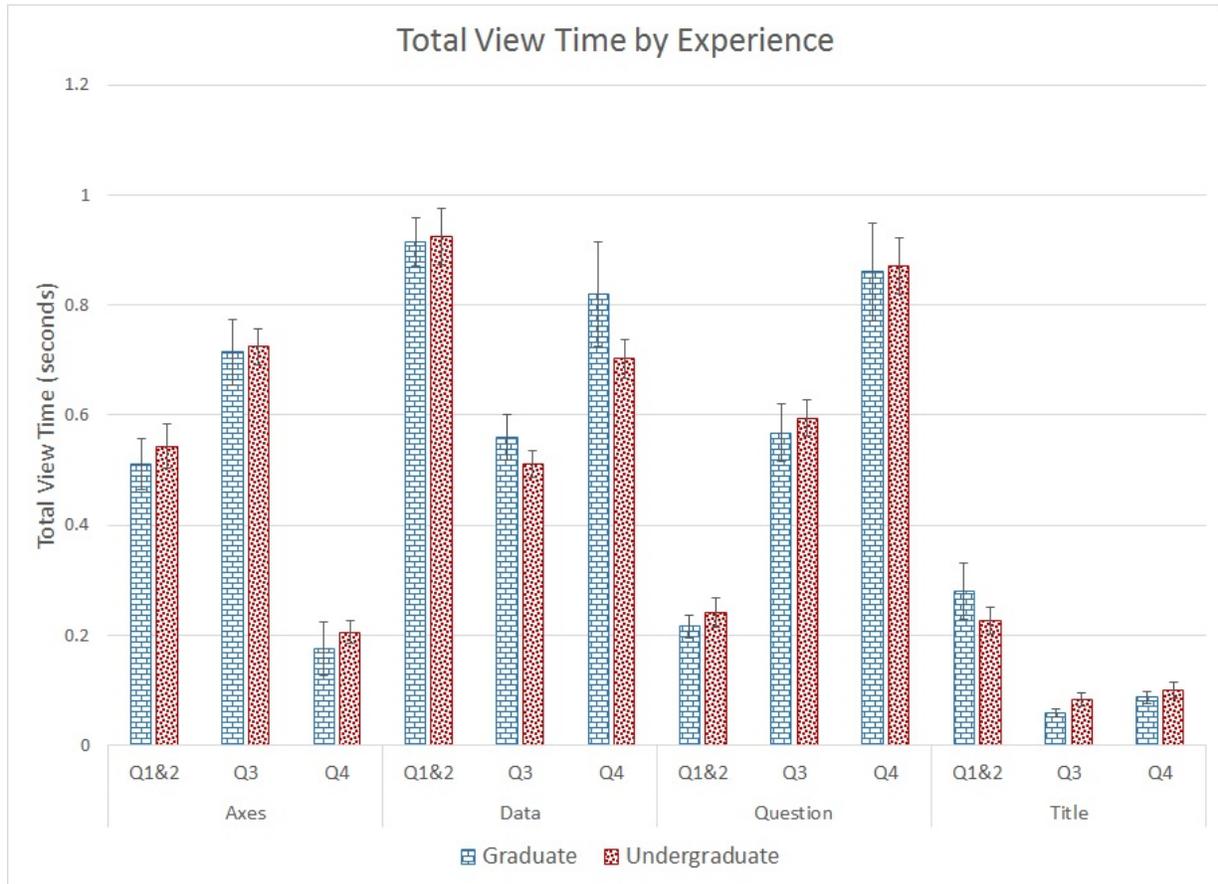
Figure 7. Graduate vs. Undergraduate normalized view time Q4. Each point represents the average normalized fraction of time spent per pixel (normalization procedure in methods) of graduate and undergraduate students for an area of interest, separated by area of interest for question 4 on all graphs.



QUALITATIVE RESULTS

Heat Maps

Figure 9 shows heat maps comparing viewing patterns of graduates and undergraduates. These heat maps support quantitative results reported above. Figure 9 shows the differences between graduates (9a) and undergraduates (9b) while viewing Q4b, an extrapolation question, of the WWTA graph. Graduates spent proportionally more attention on the data trend and undergraduates allocated more attention on the question.

Figure 8. Proportion of total view time by experience with error bars showing standard error

NVivo

In Table 2, examples 1-4 represent typical graduate and undergraduate responses grouped by graph type. Generally, graduate responses tended to be much longer, supporting the eye-tracking findings of graduates spending more of their time on most figures. When asked about the main idea of the graphs, graduate responses showed enhanced synthesis of ideas about the graph and its contents using words that were not provided in the description of the graph or its contents to describe the behavior and relationships among data more often. The majority of the undergraduate answers were very similar, if not identical to the provided title.

Examples 5-8 show population responses grouped by question type. Not only are undergraduate responses shorter, but in most cases, they also fail to include any evidence of prior knowledge or apply skills to extrapolate the data accurately. When extrapolating, graduates tended to take into consideration (or at least comment on) an increase in rate when calculating their extrapolated values. Undergraduates more often used a constant rate or failed to comment on this at all. When answering questions asking to apply prior knowledge, graduates tended to list multiple examples and elaborate more. Undergraduates mentioned single examples and included socially popular topics such as global warming and polar bears more frequently.

To determine if there was a difference in linguistics between the responses of the two populations, we conducted a text query search to determine the frequency of word use.

Figure 9. Heat maps showing attention differences between data and question/title for graduates (A) and undergraduates (B). Red indicates locations where more attention is allocated.

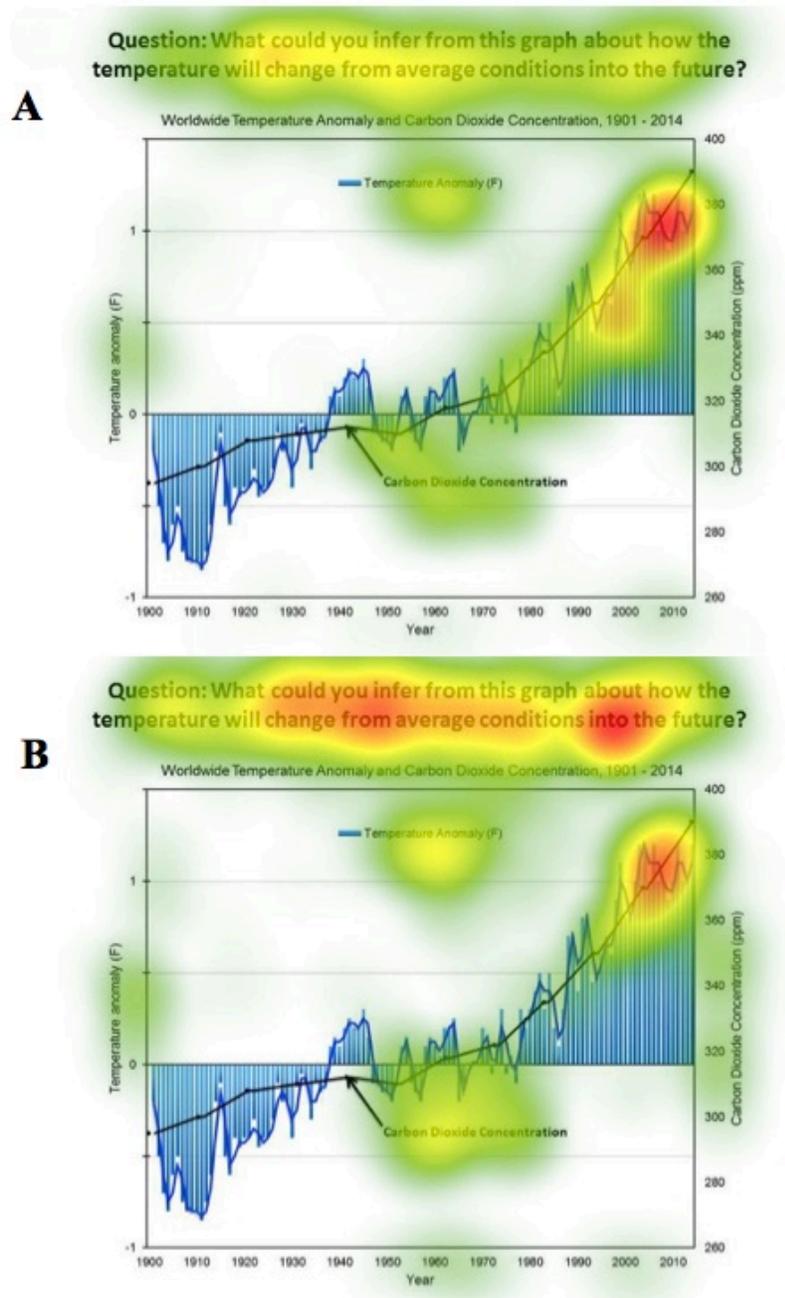
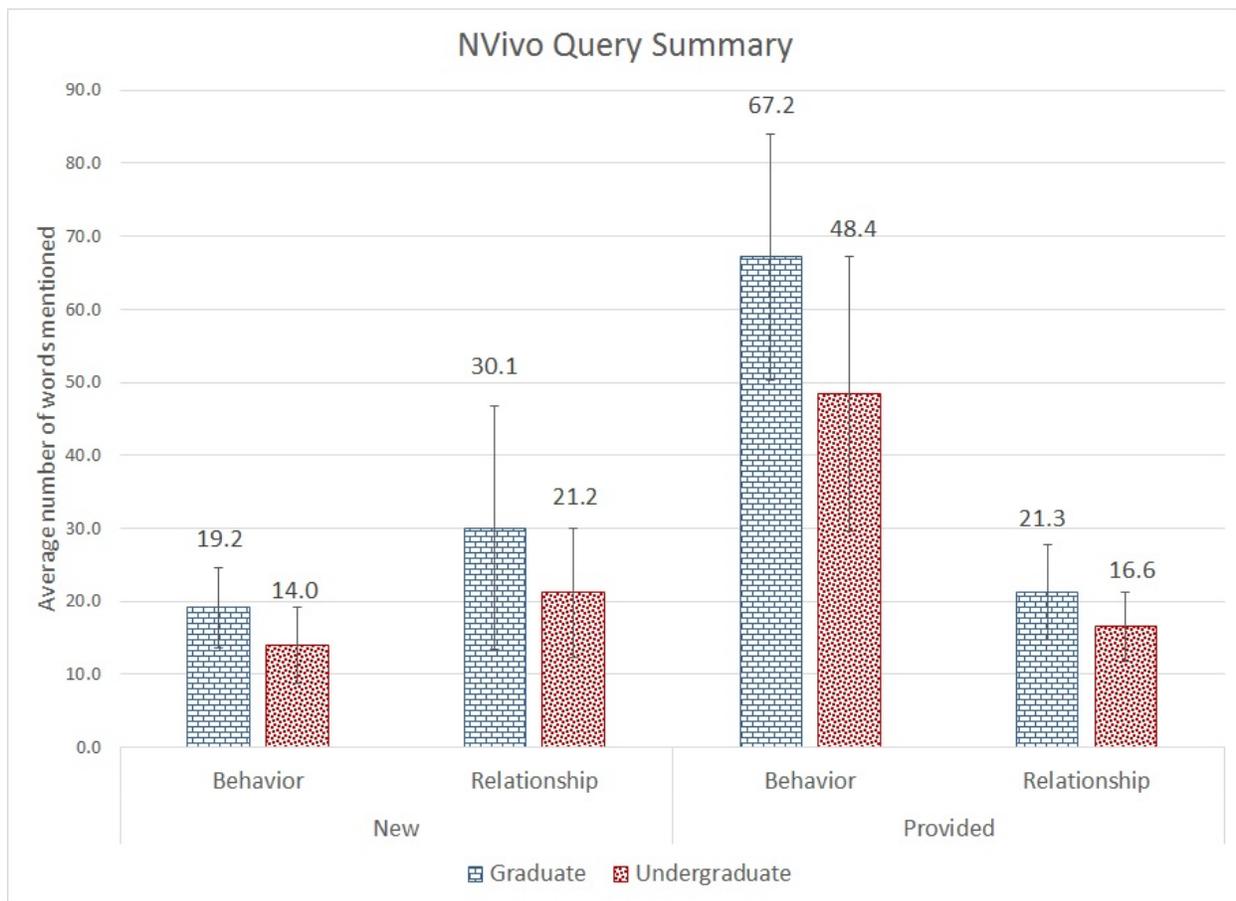


Table 2. Example participant think aloud responses. Parentheses indicate the graph the responses are describing

Graduate	Undergraduate
Question: What is the main idea of this graph? (SLGT)	
Example 1	
"To show the change in global temperature correlated with the change in sea level. You can see how the temperature has been changing globally over time as well as how the sea level has been changing over time."	"The sea level and global temperature change."
Example 2	
"This correlates how the temperature change and the sea level has generally risen from 1880 to 2010. We see that through that timeframes. We had about three degrees increase in the global temperature corresponding to about an eight-inch rise in sea level over that period of time. Pretty steady rise there."	"The main idea of this graph is that it's displaying global temperature and sea level change over time."
Question: What is the main idea of this graph? (WWTA)	
Example 3	
"It's somewhat of a combination of the previous two that we saw, in that it's showing a temperature anomaly and greenhouse gas concentration. In this case, it's focusing on carbon dioxide. Instead of just the US temperature anomaly, it's the worldwide temperature anomaly. Overall, what you see is a correlation between the two. As the CO ₂ concentration worldwide has increased, we've also seen an increase in the worldwide temperature anomaly."	"It is worldwide temperature anomaly and carbon dioxide concentration from 1901 to 2014."
Example 4	
"This graph is showing both the increase in temperature anomalies since 1900 to 2010 and the increase in carbon dioxide concentration at the same time. The implication of this graph is that the two are linked, although there's no evidence shown here to support that, other than causation -- correlation, even."	"The main idea of this graph is that it's showing carbon concentrations and temperature anomaly over time worldwide from 1901 to 2014."
Question: If the trend continues, what are some impacts that humans might see in the future? (WWTA)	
Example 5	
"Assuming the rate is flat, it's gone up two inches since 1980, so that's 20 years. Let's make it easier. From 1950 to 2010, it's gone up four inches. Assuming the same rate, then by 2050, it will have gone up to 12 inches from the base period of 1880, although I think sea level rise is increasing, so maybe slightly above 12 inches."	"Sea level change in 2050 might be 10 inches."
Example 6	
Let's take the last 50 years for an average rate, so 1960 was a little over 4 to 2010 to 8, so a little less than 4 inches. Actually, 1970 to 2010 would be 40 years, so that's about 5 inches to 8 inches, so 3 inches, so 3 more inches to 8, so 11 inches. Maybe you'd make an argument for sea level rising even more rapidly because of the way things go, so probably, it would be 11 or 12 inches of accumulative sea level change.	It looks like it's going to be probably off the graph, so above 10 inches.
Question: If the trend continues, what are some impacts that humans might see in the future? (WWTA)	
Example 7	
"Well, there are a lot of impacts to this question. Increased temperatures. We'll most likely see increase in extreme weather events, increase in precipitation as the atmosphere is able to hold more moisture, and increasing temperatures leads to more ocean acidification, which has an impact on corals and fisheries. More areas will become less inhabitable, so there will be migration issues. The sea level rise will cause some coastal areas to be uninhabitable. A lot."	"There could be a global warming crisis. Ice caps could melt. I don't think the graph tells you that, but..."
Example 8	
If there's continually increase in temperature anomalies, there could be various impacts including changing normal climate in an area. For example, there could be more droughts or other parts of the world marine or even in the continent in the United States, they might seem more rain in general as temperature's getting more and more, you get more precipitation. Also, affects most of the population in the world. In the seacoast or the sea line and as sea level rises, probably due to or definitely due to warming temperatures, then it will be affected by sea level rise.	I'd say that a lot of the ice near the North Pole and the South Pole will be melting, and the polar bears will die.

Participant responses were reviewed and recurring terms were identified. We selected a subset of these words to determine the frequency in which they were used throughout responses. Selected words are grouped into two categories, words that focus on the behavior of data (behavior) and words that focus on the relationships within and across data (relationship). Additionally, some common words included those provided by the researcher in either a question or graph title (provided), while some were newly introduced to the study by participants (new), therefore, we also included separate categories for these words. Newly introduced behavior words (new, behavior) included oscillating, constant, rapid, varies, variation, rate and exponential. Newly introduced relationship words (new, relationship) included correlate, close, normal, baseline, comparison, affect, effect and correspond. Words that were provided by the researcher relating to data behavior (provided, behavior) included trend, positive, negative, increase, decrease, change and anomaly. Words provided dealing with relationships among and across data (provided, relationship) included cause, impact, positive and negative. In addition to exact word matches (e.g. “talk”), stemmed words (e.g. “talking”) and synonyms (e.g. “speak”) were also included. All participants were included in the text search query (n=58, graduates = 13, undergraduates = 45). We found that on average, graduates used all of these words more often than undergraduates. The differences are most significant for behavior words provided by the researcher (graduates = 67.2, undergraduates = 48.4), followed by relationship words introduced by participants (graduates = 30.1, undergraduates = 21.2) (Figure 10).

Figure 10. Graphical summary of NVivo word frequency with error bars showing standard deviation



DISCUSSION

Research Question #1: How do graduate and undergraduate eye movements differ during various climate graph related tasks?

Graduates and undergraduates differ in where they allocate their attention on graphs. The magnitude of these differences appear to be dependent on the task. When given an open-ended task of identifying trends and summarizing main ideas from a graph, there is significant variance in where attention is allocated and no clear differences between the two populations is observed. When a fact-extraction question is asked, such as those in questions 3a and 3b, undergraduates spend proportionally more time on the axes, question and title. Graduates spent proportionally less time in these areas of the graph, and spent more time looking at the data elements. When the task requires individuals to extrapolate or make inferences from data provided in a graph, we see a similar, and more pronounced trend of graduates spending more of their time on the data, perhaps using it to inform their responses, and undergraduates spending more of their time reading the question, perhaps trying to determine how to approach the task.

Research Question #2: To what extent does the amount of time spent viewing various features of climate graphs differ between graduate and undergraduates?

We observed differences between graduate and undergraduate attention allocations while viewing graphs and completing graph-reading tasks. Undergraduate participants who performed highest on the pre-test displayed eye movements that were generally more graduate-like than those who scored moderately or poorly on the pre-test. Maltese et al. (2015) also found differences in performance between populations along an expertise continuum during graph reading tasks, with a strong relationship between academic year and performance. These results are also consistent with findings from Ho et al. (2013) where they found that higher prior knowledge correlated to longer view times. Okan (2016) also found that viewers with higher graph literacy spent more time on the essential elements of graphs that were needed to address the presented task.

Overall, graduates spent more total time viewing the graphs and proportionally more time on data than other elements. Conversely, undergraduates devoted a smaller percentage of their time to the data and spent more time on the axes, title and question. In a similar study with science and non-science student populations, science students tended to spend more time viewing graphs than the non-science students (Yen et al., 2012). Additionally, an expert-expert study by Roth & Bowen (2010) demonstrated that when the expert was familiar with the content displayed in the graph, they were able to focus more immediately on the task, which may explain why graduate students in this study viewed the data elements in the graph proportionally longer than undergraduates. Furthermore, graduates in this study could use their previous knowledge of and familiarity with graphs to spend more time on the data features of the graph that were important to solve the task presented to them, where undergraduates needed to gain familiarity with the non-essential elements as well as the essential elements to address the tasks.

Research Question #3: To what extent do graduate and undergraduate verbal interpretations of climate graphs differ?

The quantitative findings were supported by participant verbal responses. Graduate responses tended to be longer than undergraduate explanations. Longer responses to these questions could be attributed to the fact that graduates had more prior knowledge about climate change and could therefore better attend to these questions. Additionally, undergraduates more often described general trends, while graduates identified more specific patterns. Graduate responses were longer and more descriptive, using more vivid language to describe the behavior of data (i.e. oscillating, rate, exponential, etc.) as well as the relationships among and between data (i.e. correlate, comparison, correspond, etc.). More education and experience with graphs could explain this finding, allowing graduate students to draw on information stored in long-term memory to make high-level interpretations and inferences of the information provided in the displays.

Table 3 highlights some of the largest differences we saw between graduate and undergraduate students, specifically within the Data AOI. Not only were the graduate pre-test average scores higher, they also allocated larger proportions of their attention to the data on all graphs for all questions. These differences were largest for questions 4a and 4b that asked participants to explain or make inferences based on information not provided on the graphs. The verbal

responses indicated that this allocation of more attention to data was used to provide more descriptive responses about the graph, integrating the provided graph with their inference or prediction about the future.

Table 3. Summary of pre-test, Data AOI and verbal response comparisons between graduates and undergraduates.

Pre-test score		Graduates	Undergraduates
		81.22%	69.81%
Proportion of view time on <i>Data AOI</i>	Q1&2	0.50	0.49
	Q3	0.39	0.36
	Q4	0.46	0.37
Average number of <i>New</i> words mentioned	Behavior	19.20	14.00
	Relationship	30.10	21.20
Average number of <i>Provided</i> words mentioned	Behavior	67.20	48.40
	Relationship	21.30	16.60

CONCLUSIONS

Overall, this exploratory research adds to the limited literature that has employed eye-tracking to better understand how climate change graphs are interpreted by graduate and more novice undergraduate populations. Our findings indicate that graduate and undergraduate students spend different amounts of time on graph features where the type of task presented to a viewer can accentuate these differences. The findings also add to work that has successfully employed eye-tracking to understand graph comprehension in general (Kim, Lombardino, Cowles, & Altmann, 2014; Okan et al., 2016; Woller-Carter, Okan, Cokely, & Garcia-Retamero, 2012) where we have incorporated graphs with higher complexity than most previous studies.

When designing climate graphs for novice populations, developers should be aware that novices do not have the same prior knowledge or experience with graphs (or climate change content) as experts. The results of this research show that these factors can affect novice view times on the most salient graphical elements. As such, design considerations should allow the most important aspects of the graph to be readily accessible to novice viewers (Harold et al., 2016). The use of graphical animations or callouts could be a strategy to address such issues, however a variety of design issues need to be considered with the target audience in mind before implementation (Kriz & Hegarty, 2007).

Graph interpretation becomes easier with practice reading and interpreting complicated graphs (Freedman & Shah, 2002), which suggests that novices should use graphs more often to interpret data to become more expert-like. Implications for teaching and learning from our results suggest that it would be beneficial to emphasize the importance of novice attention to data and data extraction techniques where instructors should offer scaffolding and training aimed to help students direct attention to data elements (i.e. axes, legends, data trends, etc.) (Curcio, 1987; Hao Wang et al., 2012). In order to direct proportionally more attention toward data and interpretation, it could be beneficial for students to formulate their own title, rather than being provided one, encouraging them to assimilate their own ideas about what information the graph is conveying and allowing more viewing time on the essential data elements of the graphical display.

Due to the exploratory nature of this study, our work has some limitations, some of which could be explored in our suggestions for future work. Our first limitation was that the concurrent verbal protocol used in this study may have influenced view times, which could attribute to graduates taking longer to answer the questions in general but it is unlikely that viewers all looked at the same graph elements when talking, so our results of graduates looking at data elements longer than undergraduates is likely still valid. While this study tested for graph reading skill, it did not assess climate literacy and prior climate knowledge. Additionally, the lack of statistical significance with some of the eye-tracking metrics used in the study could be due to the exploratory nature of this study and the typical small sample sizes characteristic of most eye-tracking studies, including this one. This limitation could, in part, be addressed in future work by using more focused stimuli and questions to decrease the variance in participant view times, and increase the likelihood of detecting differences in the data. Furthermore, the influence of worldview and political affiliation were outside of the scope of the current study and therefore not explored. Finally, this study was quasi-experimental in design where our population was a sample of convenience from a large-enrollment introductory science course at a major research university in the southeastern United States and local graduate students in residence at the same university. As such, our results may not be transferrable to other populations or settings. Despite the

limitations of this work, this study is one of few published eye-tracking studies in the geosciences aiming to understand how students in various locations on the expertise continuum attend to climate graphs. We hope that others in the field can expand upon this exploratory work to improve the teaching and learning process surrounding scientific graph reading among students as well as use it to study how to design climate graphs that are effective in eliciting viewer attention and understanding.

Suggestions for future work include an exploration of the following as they affect performance; the influence of graph type (including bar versus line graphs), the combination of bar and line graphs, and the effect of graph complexity; the effect of graph content (e.g., climate change versus other topics); the influence of using a concurrent versus retrospective verbal protocol in the eye-tracking study design; and the influence of the type of climate content covered (i.e. greenhouse gases, sea level rise, etc.), as some concepts may draw more heavily on prior knowledge or misconceptions than others. Furthermore, future research may focus on the influence of climate literacy, graph modifications, animations and addition of supplementary text on the interpretation and reading of climate graphs. The effect of graph reading ability and climate literacy on climate graph reading could also be expanded to a variety of populations (i.e., pre-service teachers, K-12 students, the public). Finally, future studies may choose to explore the influence of worldview and political affiliation on respondent view time and attention.

ACKNOWLEDGMENTS

This research was funded in part by EarthLabs Climate (NSF-DRK12) grants DUE-1019721, DUE-1019703, DUE-1019815. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the view of the National Science Foundation. We would also like to thank Anne Gold for helping to develop the pre-test for this study, Sarah Luginbuhl for locating and modifying graph images and the Southeast Climate Adaptation Science Center for their funding and support through the Global Change Graduate Fellows Program.

AUTHOR BIOGRAPHIES

Rachel Atkins is a PhD student in the Marine, Earth and Atmospheric Sciences Department at NC State University. Her research has centered on student learning in the geoscience investigating spatial skill evolution, climate communication, and the application of Augmented Reality in the undergraduate classroom.

Karen McNeal is Associate Professor of Geoscience Education and Geocognition in the Department of Geosciences at Auburn University. Her research is focused on affect and cognition in learning about the geosciences through qualitative and quantitative methods.

REFERENCES

- Aksit, O., McNeal, K. S., Gold, A. U., Libarkin, J. C., & Harris, S. (2017). The influence of instruction, prior knowledge, and values on climate change risk perception among undergraduates. *Journal of Research in Science Teaching*, (October 2017), 550–572. <https://doi.org/10.1002/tea.21430>
- Beattie, G., & McGuire, L. (2012). See no evil? Only implicit attitudes predict unconscious eye movements towards images of climate change. *Semiotica*, 2012(192), 315–339. <https://doi.org/10.1515/sem-2012-0066>
- Bojko, A. (2013). *Eye tracking the user experience: A practical guide to research*. Rosenfeld Media. Retrieved from <http://rosenfeldmedia.com/books/eye-tracking-the-user-experience/>
- Bostrom, A., Bohm, G., & O'Connor, R. E. (2013). Targeting and tailoring climate change communications. *Wiley Interdisciplinary Reviews: Climate Change*, 4(5), 447–455. <https://doi.org/10.1002/wcc.234>
- Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, 4(2), 75–100. <https://doi.org/10.1037/1076-898X.4.2.75>
- Cook, J., Oreskes, N., Doran, P. T., Reusswig, F., Carlton, J. S., Perry-hill, R., ... Richardson, M. (2007). The influence of political ideology on trust in science Quantifying the consensus on anthropogenic global warming in the scientific literature. *Environmental Research Letters*, 8(2), 24024. <https://doi.org/10.1088/1748-9326/8/2/024024>
- Curcio, F. R. (1987). Comprehension of Mathematical Relationships Expressed in Graphs. *Journal for Research in Mathematics Education*, 18(5), 382. <https://doi.org/10.2307/749086>
- Daly, J. C., & Cohen, J. (1978). Statistical Power Analysis for the Behavioral Sciences, Revised Edition. *Journal of the American Statistical Association*, 73(363), 680. <https://doi.org/10.2307/2286629>

- Fabrikant, S. I., Hespanha, S. R., & Hegarty, M. (2010). Cognitively Inspired and Perceptually Salient Graphic Displays for Efficient Spatial Inference Making. *Annals of the Association of American Geographers*, 100(1), 13–29. <https://doi.org/10.1080/00045600903362378>
- Freedman, E. G., & Shah, P. (2002). Toward a Model of Knowledge-Based Graph Comprehension. *Diagrams*, 18–30. https://doi.org/10.1007/3-540-46037-3_3
- Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and institutional implications. *Journal for Research in Mathematics Education*, 32(2), 124–158. <https://doi.org/10.2307/749671>
- Gautier, C., Deutsch, K., & Rebach, S. (2006). Misconceptions about the Greenhouse Effect. *Journal of Geoscience Education*, 54(May), 386–395.
- Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Studies in Science Education*. <https://doi.org/10.1080/03057267.2011.605307>
- Hao Wang, Z., Wei, S., Ding, W., Chen, X., Wang, X., & Hu, K. (2012). Students' cognitive reasoning of graphs: Characteristics and progression. *International Journal of Science Education*, 34(13), 2015–2041. <https://doi.org/10.1080/09500693.2012.709333>
- Harold, J., Lorenzoni, I., Shipley, T. F., & Coventry, K. R. (2016, December 1). Cognitive and psychological science insights to improve climate change data visualization. *Nature Climate Change*. Nature Publishing Group. <https://doi.org/10.1038/nclimate3162>
- Ho, H., Tsai, M.-J., Wang, C.-Y., & Tsai, C.-C. (2014). Prior knowledge and online inquiry-based science reading: Evidence from eye tracking. *International Journal of Science & Mathematics Education*, 12(3), 525–554. <https://doi.org/10.1007/s10763-013-9489-6>
- Huang, W., Eades, P., & Hong, S.-H. (2009). Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization*, 8(3), 139–152. <https://doi.org/10.1057/ivs.2009.10>
- IPCC (United Nations). (2013). *Fifth Assessment Report - Climate Change 2013*. IPCC. Retrieved from <https://www.ipcc.ch/report/ar5/wg1/>
- Kahan, D. M., Jenkins-Smith, H., & Braman, D. (2011). Cultural cognition of scientific consensus. *Journal of Risk Research*, 14(2), 147–174. <https://doi.org/10.1080/13669877.2010.511246>
- Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, 2(10), 732–735. <https://doi.org/10.1038/nclimate1547>
- Kastens, K. A., Shipley, T. F., & Boone, A. P. (2016). What geoscience experts and novices look at, and what they see, when viewing data visualizations. *Journal of Astronomy & Earth Sciences Education*, 3(1).
- Kim, S., Lombardino, L. J., Cowles, W., & Altmann, L. J. (2014). Investigating graph comprehension in students with dyslexia: An eye tracking study. *Research in Developmental Disabilities*, 35(7), 1609–1622. <https://doi.org/10.1016/j.ridd.2014.03.043>
- Korner, C. (2004). Sequential processing in comprehension of hierarchical graphs. *Applied Cognitive Psychology*, 18(4), 467–480. <https://doi.org/10.1002/acp.997>
- Korner, C. (2011). Eye movements reveal distinct search and reasoning processes in comprehension of complex graphs. *Applied Cognitive Psychology*, 25(6), 893–905. <https://doi.org/10.1002/acp.1766>
- Kriz, S., & Hegarty, M. (2007). Top-down and bottom-up influences on learning from animations. *International Journal of Human Computer Studies*, 65(11), 911–930. <https://doi.org/10.1016/j.ijhcs.2007.06.005>
- Land, M. F. (2012). Oculomotor behaviour in vertebrates and invertebrates. In *The Oxford Handbook of Eye Movements* (pp. 3–15). Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199539789.013.0001>
- Lee, T. M., Markowitz, E. M., Howe, P. D., Ko, C.-Y., & Leiserowitz, A. A. (2015). Predictors of public climate change awareness and risk perception around the world. *Nature Climate Change*, 5(11), 1014–1020. <https://doi.org/10.1038/nclimate2728>
- Leiserowitz, A. A. (2005). American risk perceptions: Is climate change dangerous? *Risk Analysis*. <https://doi.org/10.1111/j.1540-6261.2005.00690.x>
- Leiserowitz, A. A., Smith, N., & Marlon, J. R. (2010). Americans' knowledge of climate change. *Yale Project on Climate Change Communication*, 1–60. Retrieved from <http://environment.yale.edu/climate/files/ClimateChangeKnowledge2010.pdf%5Cnpapers3://publication/uuid/230CA706-26C2-4156-8133-1AC9E7701E6A>
- Libarkin, J. C., Kurdziel, J. P., & Anderson, S. W. (2007). College student conceptions of geological time and the disconnect between ordering and scale. *Journal of Geoscience Education*, 55(5). Retrieved from <http://search.proquest.com/docview/202781283/fulltextPDF/ED86C75E2DCC461BPQ/1?accountid=12725>
- Libarkin, J. C., Thomas, S. R., & Ordning, G. (2015). Factor analysis of drawings: Application to college student models of the greenhouse effect. *International Journal of Science Education*, 37(13), 2214–2236. <https://doi.org/10.1080/09500693.2015.1074757>
- Lombardi, D., & Sinatra, G. M. (2012). College Students' perceptions about the plausibility of human-induced climate change. *Research in Science Education*, 42(2), 201–217. <https://doi.org/10.1007/s11165-010-9196-z>

- Maltese, A. V., Harsh, J. A., & Svetina, D. (2015). Data visualization literacy: Investigating data interpretation along the novice--expert continuum. *Journal of College Science Teaching*, 45(1), 83–83. https://doi.org/10.2505/4/jcst15_045_01_84
- McNeal, K. S., Libarkin, J. C., Ledley, T. S., Bardar, E., Haddad, N., Ellins, K., & Dutta, S. (2014). The Role of research in online curriculum development: The case of earthlabs climate change and earth system modules. *Journal of Geoscience Education*, 62(4), 560–577. <https://doi.org/10.5408/13-060.1>
- McNeal, K. S., Spry, J. M., Mitra, R., & Tipton, J. L. (2014). Measuring student engagement, knowledge, and perceptions of climate change in an introductory environmental geology course. *Journal of Geoscience Education*, 62(4), 655–667. <https://doi.org/10.5408/13-111.1>
- McNeal, K. S., Walker, S. L., & Rutherford, D. (2014). Assessment of 6- to 20-grade educators' climate knowledge and perceptions: Results from the climate stewardship survey. *Journal of Geoscience Education*, 62(4), 645–654. <https://doi.org/10.5408/13-098.1>
- Mitra, R., McNeal, K. S., & Bondell, H. D. (2017). Pupillary response to complex interdependent tasks: A cognitive-load theory perspective. *Behavior Research Methods*, 49(5), 1905–1919. <https://doi.org/10.3758/s13428-016-0833-y>
- Muldner, K., Christopherson, R., Atkinson, R., & Burlison, W. (2009). Investigating the utility of eye-tracking. *User Modeling Adaptation and Personalization*, 31, 138–149. https://doi.org/10.1007/978-3-642-02247-0_15
- National Oceanic and Atmospheric Administration (NOAA). (2009). *Climate literacy: The essential principles of climate sciences*. Science. Retrieved from http://oceanservice.noaa.gov/education/literacy/climate_literacy.pdf
- Niebert, K., & Gropengiesser, H. (2012). Understanding and communicating climate change in metaphors. *Environmental Education Research*, 19(3), 282–302. <https://doi.org/10.1080/13504622.2012.690855>
- O'Neill, S. J., Boykoff, M., Niemeyer, S., & Day, S. A. (2013). On the use of imagery for climate change engagement. *Global Environmental Change*, 23(2), 413–421. <https://doi.org/10.1016/j.gloenvcha.2012.11.006>
- Okan, Y., Galesic, M., & Garcia-Retamero, R. (2016). How people with low and high graph literacy process health graphs: Evidence from eye-tracking. *Journal of Behavioral Decision Making*, 29(2–3), 271–294. <https://doi.org/10.1002/bdm.1891>
- Ramat, S., Leigh, R. J., Zee, D. S., Shaikh, A. G., & Optican, L. M. (2008). Applying saccade models to account for oscillations. *Progress in Brain Research*. [https://doi.org/10.1016/S0079-6123\(08\)00616-X](https://doi.org/10.1016/S0079-6123(08)00616-X)
- Ratwani, R. M., & Trafton, G. J. (2008). Shedding light on the graph schema: Perceptual features versus invariant structure. *Psychonomic Bulletin & Review*, 15(4), 757–62. <https://doi.org/10.3758/pbr.15.4.757>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422. <https://doi.org/10.1037/0033-2909.124.3.372>
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506. <https://doi.org/10.1080/17470210902816461>
- Rebich, S., & Gautier, C. (2005). Concept mapping to reveal prior knowledge and conceptual change in a mock summit course on global climate change. *Journal of Geoscience Education*, 53(4), 355–365. Retrieved from <http://search.proquest.com/docview/202779335/fulltextPDF/FB28B5A24C0D4D8DPQ/1?accountid=12725>
- Resnick, I., Kastens, K. A., & Shipley, T. F. (2018). How students reason about visualizations from large professionally collected data sets: A study of students approaching the threshold of data proficiency. *Journal of Geoscience Education*, 66(1), 55–76. <https://doi.org/10.1080/10899995.2018.1411724>
- Roth, W.-M., & Bowen, G. M. (2010). When are graphs worth ten thousand words? An expert-expert study. *Cognition and Instruction*, 21(4), 429–473. <https://doi.org/10.1207/s1532690xci2104>
- Rule, A. C., & Meyer, M. A. (2009). Teaching urban high school students global climate change information and graph interpretation skills using evidence from the scientific literature. *Journal of Geoscience Education*, 57(5), 335–347. <https://doi.org/10.5408/1.3559674>
- Shah, P., & Freedman, E. G. (2011). Bar and line graph comprehension: An interaction of top-down and bottom-up processes. *Topics in Cognitive Science*, 3(3), 560–578. <https://doi.org/10.1111/j.1756-8765.2009.01066.x>
- Shah, P., Freedman, E. G., & Vekiri, I. (2006). The comprehension of quantitative information in graphical displays. *The Cambridge Handbook of Visuospatial Thinking*, 426–476. <https://doi.org/10.1017/CBO9780511610448.012>
- Shepardson, D. P. (2011). Seventh grade students' mental models of the greenhouse effect. *Environmental Education Research*, 17(1), 1–17. <https://doi.org/10.1080/13504620903564549>
- Somerville, R. C. J., & Hassol, S. J. (2011). Communicating the science of climate change. *Physics Today Physics Today Physics Today*, 64(64), 48–39. <https://doi.org/10.1063/PT.3.1295>
- Sterman, J. D., & Sweeney, L. B. (2002). Cloudy skies: Assessing public understanding of global warming. *System Dynamics Review*, 18(2), 207–240. Retrieved from <http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=1&sid=6eacc440-8429-471d-a86d-4893ed7569f0%40sessionmgr101>
- Trafton, G. J., Kirschenbaum, S. S., Tsui, T. L., Miyamoto, R. T., Ballas, J. A., & Raymond, P. D. (2000). Turning pictures into numbers: Extracting and generating information from complex visualizations. *International Journal of Human-Computer Studies*, 53(5), 827–850. <https://doi.org/10.1006/ijhc.2000.0419>
- Trickett, S. B., & Trafton, G. J. (2006). Toward a comprehensive model of graph comprehension: Making the case for spatial cognition. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 4045 LNAI, pp. 286–300). Springer, Berlin, Heidelberg. https://doi.org/10.1007/11783183_38

- U.S. Environmental Protection Agency (2014). *Climate Change Indicators in the United States, 2014*. Retrieved from www3.epa.gov/climatechange/science/
- U.S. Global Change Research Program (USGCRP). (2016). Understand global change. Retrieved March 1, 2016, from <http://www.globalchange.gov/climate-change>
- van der Linden, S. L., Leiserowitz, A. A., Feinberg, G. D., & Maibach, E. W. (2014). How to communicate the scientific consensus on climate change: plain facts, pie charts or metaphors? *Climatic Change*, *126*(1–2), 255–262. <https://doi.org/10.1007/s10584-014-1190-4>
- van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology*, *19*(2), 205–221. <https://doi.org/10.1002/acp.1112>
- van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction*, *20*(2), 95–99. <https://doi.org/10.1016/j.learninstruc.2009.02.009>
- Weber, E. U., & Stern, P. C. (2011). Public understanding of climate change in the United States. *American Psychologist*, *66*(4), 315–328. <https://doi.org/10.1037/a0023253>
- Winn, B. (1987). Charts, graphs, and diagrams in educational materials. *The Psychology of Illustrations: Vol 1 Basic Research*, 152–198.
- Woller-Carter, M. M., Okan, Y., Cokely, E. T., & Garcia-Retamero, R. (2012). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting: Communicating and distorting risks with graphs: An eye-tracking study*. *56*(1), 1723–1727. <https://doi.org/10.1177/1071181312561345>
- Yen, M. H., Lee, C. N., & Yang, Y. C. (2012). Eye movement patterns in solving scientific graph problems. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 7352 LNAI, pp. 343–345). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-31223-6_46

APPENDICES MATERIALS

Graph Reading Pre-Test

The pre-test was provided to all participants in this study.

Graph Proficiency Questionnaire

1) Your name: _____

2) How *confident* are you...

	Very Confident	Somewhat confident	Somewhat unconfident	Very unconfident
... with interpreting scientific graphs?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

3) How *good* are you at...

	Novice	Advanced beginner	Competent	Proficient	Expert
...interpreting scientific graphs?	<input type="radio"/>				

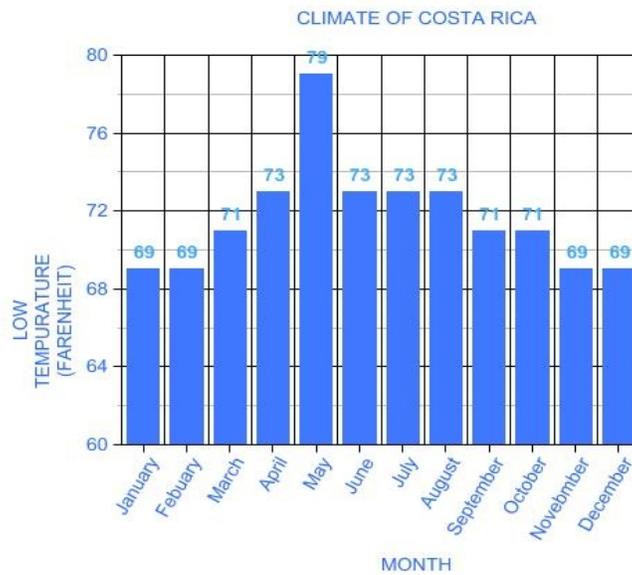
4) How often do you...

	Never	Once a Year or Less	Several Times a Year	Once a Month	2-3 Times a Month	Once a Week	2-3 Times a Week	Daily
...construct graphs using paper and pencil?	<input type="radio"/>							
...interpret existing graphs?	<input type="radio"/>							
...use computers for data analysis?	<input type="radio"/>							
...use computers for graphing data?	<input type="radio"/>							
...use Excel for data analysis?	<input type="radio"/>							
...use Excel for graphing data?	<input type="radio"/>							
...do other graphing or graph reading tasks?	<input type="radio"/>							
Please describe other graphing or graph reading tasks:								

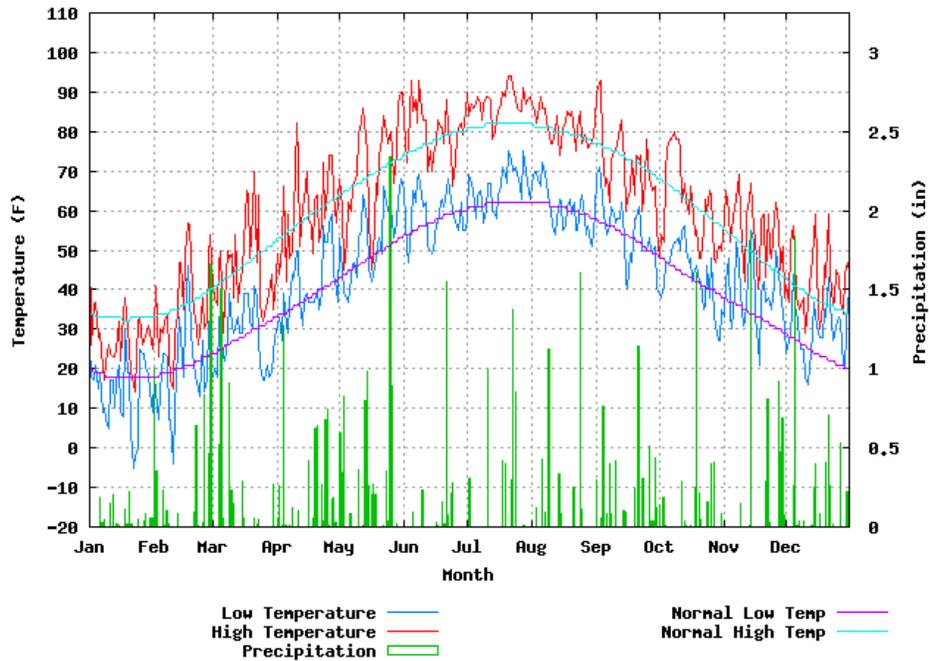
5) How often do you...

	Never	Once a Year or Less	Several Times a Year	Once a Month	2-3 Times a Month	Once a Week	2-3 Times a Week	Daily
...read Wikipedia articles about a science topic?	<input type="radio"/>							
...read internet articles or blogs about a science topic?	<input type="radio"/>							
...read science magazines?	<input type="radio"/>							
...read science research articles?	<input type="radio"/>							
...read science text books?	<input type="radio"/>							
...do other science reading?	<input type="radio"/>							
Please describe other science reading:								

6) Circle the y-axis or axes on the following graph:



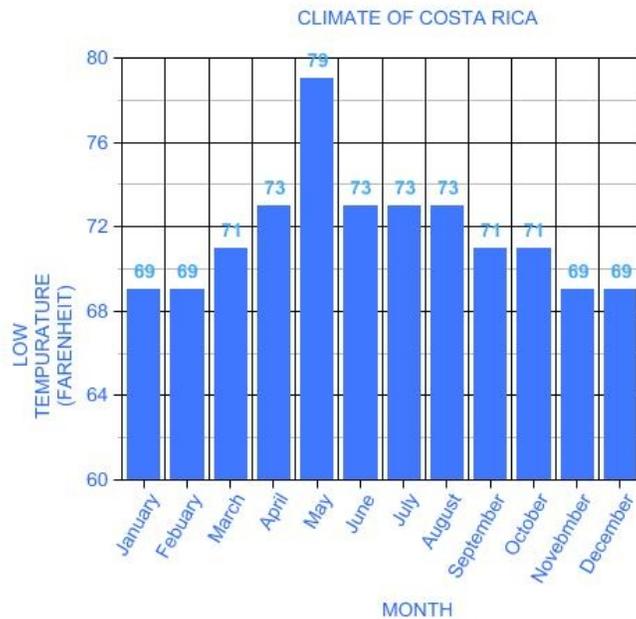
7) Circle the x-axis or axes on the following graph:



8) What is an independent variable?

- I am not sure.
- _____

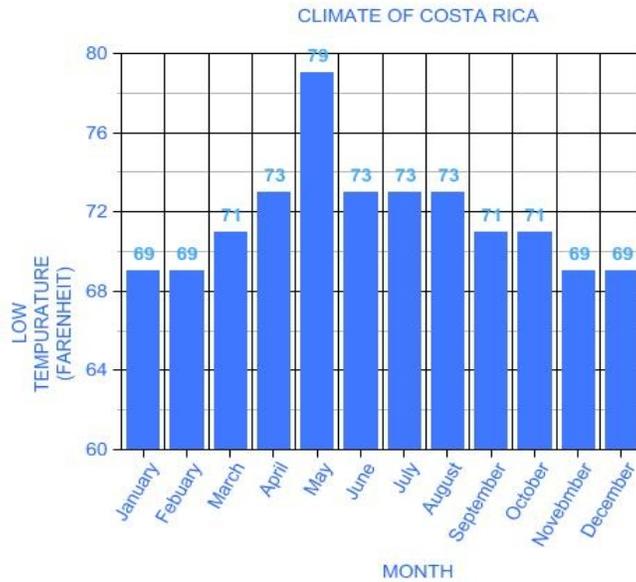
9) Circle the independent variable on the following graph:



10) What is a dependent variable?

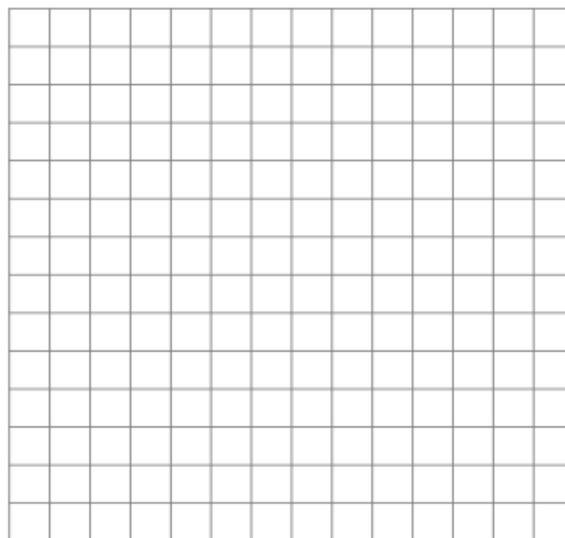
- I am not sure.
- _____

11) Circle the dependent variable on the following graph:



12) Please graph the participant data in the table below from the hypothetical online MEA 100 section. Make sure the graph is complete.

	2012	2013	2014	2015
Participants in online MEA 100	243	251	193	355



13) Please graph the student data in the table below from a previous study. Make sure the graph is complete.

	2014			2015		
	Left-handed	Right-handed	Ambidextrous	Left-handed	Right-handed	Ambidextrous
Number of Students	5	1	1	3	4	3



14) What did you find most difficult in completing the graphing tasks above?