Journal of Eye Movement Research
7(5):2, 1-15

Comparing Experts and Novices on Scaffolded Data Visualizations using Eye-tracking

Kathryn Stofer
University of Florida

Xuan Che
National Institutes of Health

Spatially-based scientific data visualizations are becoming widely available, yet they are often not optimized for novice audiences. This study follows after an investigation of expert and novice meaning-making from scaffolded data visualizations using clinical interviews. Using eye-tracking and concurrent interviewing, we examined quantitative fixation and AOI data and qualitative scan path data for two expertise groups (N = 20) on five versions of scaffolded global ocean data visualizations. We found influences of expertise, scaffolding, and trial. In accordance with our clinical interview findings, experts use different meaning-making strategies from novices, but novice performance improves with scaffolding and guided practice, providing triangulation. Eye-tracking data also provide insight on meaning-making and effectiveness of scaffolding that clinical interviews alone did not.

Keywords: Scientific data visualizations, expert-novice, scaffolding, meaning-making, eye-tracking

Introduction

Educators and neuroscientists both have been hoping to find ways to integrate the more applied studies of education tasks with the more basic studies of neuroscience, with limited success to this point (Bransford, Brown, & Cocking, 2000; Bruer, 2006; Daniel, 2012). However, for years these two traditions have been examining tasks at different scales (Jensen, 1998). Traditionally education studies take place in classrooms and investigate higher-order tasks such as reading and comprehending paragraphs or even whole curricula, in what Jensen calls "action research" (1998, p. 5). On the other hand neuroscience studies take place in laboratories and tend to focus on cellular-level functioning or smaller components of tasks, such as letter recognition. Researchers are beginning to integrate functional magnetic resonance imaging and more traditional psychology tasks (Lobben, Lawrence, & Olson, 2009; Lobben, Olson, & Huang, 2005), studying the same task at the same level but with two different methodologies, thus comparing the behavioral and neural responses directly.

Expert performance has been examined in a number of cognitive domains, including chess (Chase & Simon, 1973; Reingold, Charness, Pomplun, & Stampe, 2001; Sheridan & Reingold, 2014), medicine (Benner, 1982; Gegenfurtner, Siewiorek, Lehtinen, & Säljö, 2013; Schubert, Denmark, Crandall, Grome, & Pappas, 2013), map perception (Anderson & Leinhardt, 2002; Kalyuga, Ayres, Chandler, & Sweller, 2003; Ooms, De Maeyer, & Fack, 2014; Ooms, De Maeyer, Fack, Van Assche, & Witlox, 2012), scientific observation (Bransford et al., 2000; Eberbach & Crowley, 2009; Kastens, Agrawal, & Liben, 2009), learning (Ertmer & Newby, 1996; Jee, Uttal, Spiegel, & Diamond, 2013; Walsh et al., 2011), and physics (Chi, Feltovich, & Glaser, 1981; Elby, 2001; Larkin, McDermott, Simon, & Simon, 1980; Rottman, Gentner, & Goldwater, 2012; Wolf, 2012), to name just a few. One must first understand expert task performance and how it differs from novice performance in order to teach novices how to solve tasks (Edelson & Gordin, 1997, 1998; Middendorf & Pace, 2004). Acquiring expertise in a domain requires deliberate practice (Ericsson, Krampe, & Tesch-Römer, 1993) and results in different mental representations in experts and novices (Ericsson & Lehmann, 1996; Gauthier & Tarr, 2002).

DOI 10.16910/jemr.7.5.2

ISSN 1995-8692

Eye-tracking is starting to bridge the gap between neuroscience and education. It was originally used for map reading in the 1970's (Krygier & Wood, 2011; Steinke, 1987), but was used less frequently after the 1980's (Çöltekin, Heil, Garlandini, & Fabrikant, 2009). Recently, groups have returned to investigating eye movements on a variety of higher-order, real-world tasks such as map reading (Canham & Hegarty, 2010; Hegarty, 2013), data visualization interpretation (Libarkin, Clark, & Simmon, 2010; Steffke & Libarkin, 2012, 2013), problem solving (Grant & Spivey, 2003; van Gog, 2006), real-world scene investigation (Henderson, Brockmole, Castelhano, & Mack, 2007), and even museum exhibit interaction (Filippini-Fantoni, Jaebker, Bauer, & Stofer, 2013). While some of these have been validated in comparison to existing accepted neuroscience methodologies (Holmqvist et al., 2011), few if any studies have yet provided a direct comparison with a traditionally educational method.

Eye-tracking in particular has been used to investigate novice-expert differences in unconscious strategies for map reading. Experts and others with higher prior knowledge do show differences in number and length of fixations on a variety of map interpretation tasks (Canham & Hegarty, 2010; Çöltekin, Fabrikant, & Lacayo, 2010; Çöltekin et al., 2009; Ooms et al., 2012). However, Hegarty (2013) calls for more empirical data to support map and visualization design.

While several studies have explored scaffolding internal to complex graphics and visualizations and shown various improvements with novice populations (Canham & Hegarty, 2010; Libarkin, Thomas, & Ruetenik, 2013; Phipps & Rowe, 2010), global satellite visualizations without additional symbols are a novel modality for this empirical exploration, especially with eye-tracking. Previous work has demonstrated that unchanged scientific versions of similar global visualizations are opaque to non-specialist visitors (Haley Goldman, Kessler, & Danter, 2010; Phipps & Rowe, 2010; Rowe, Stofer, Barthel, & Hunter, 2010). These visualizations are akin to those used on spherical display platforms in public educational settings, such as the Science on a Sphere ®, deployed in nearly 100 museums worldwide (National Oceanic and Atmospheric Administration, n.d.).

Therefore, this study is part of a larger project to examine performance on a real-world task of making meaning from data visualizations, using both clinical interviewing and eye-tracking to compare experts and novices. Here, we use eye-tracking to determine whether experts and novices show different patterns of gaze on visualizations they report understanding differently. Additionally, we wanted to determine whether scaffolding changed the viewing patterns of either expert or novice subjects.

Methods

This study employs a between- and within-subjects design to balance experimental control and ecological validity, as suggested by Çöltekin et al (2009). Subjects had previously participated in clinical interviews where they viewed visualizations related to those in the eye-tracking experiment here. See Stofer (2013) for full description of the clinical interviews. Ten novices and eight experts were invited at random from the clinical interview subjects to participate in the eye-tracking experiment.

Novices were undergraduates at a large research university in the U.S. Pacific Northwest in their first two years of study who were not pursuing a natural science or engineering major, recruited by flyers in the community. Experts were professional researchers at the same institution who had earned a Ph.D. in oceanography and had at least five years of experience beyond the Ph.D. Experts were recruited by randomly sampling an alphabetical list of the university's oceanography department professors who met the qualifications. No subject was color-blind.

Stimuli for the experiment were the same as in Stofer (2013). Stimuli were 800 x 600 pixel versions of a single global satellite data visualization with different levels of scaffolding (Wood, Bruner, & Ross, 1976), intended to bring the two groups closer together in meaning-making. Five versions were created for each of three topics: Sea Surface Temperature, Sea Surface Temperature Anomaly, and Chlorophyll Concentration. Sea Surface Temperature and Chlorophyll Concentration present continuous data while Sea Surface Temperature Anomaly represents a diverging dataset. Scaffolds were chosen based on previous work (Phipps & Rowe, 2010; Rowe et al., 2010). In the interview, each subject saw stimuli from two topics, up to five versions of each.

Four types of scaffolding were applied to each of the three topics; along with the unscaffolded (US) version, a

DOI 10.16910/jemr.7.5.2

ISSN 1995-8692

total of five visualizations were shown to each subject. Geographic scaffolding (GS), involved adding labels for five continents and three ocean basins. Color scaffolding (CS) meant that the six-hued "rainbow" color ramp was changed to a continuous, single-hued, varying brightness ramp for Sea Surface Temperature and Chlorophyll Concentration and a diverging, two-hued varying brightness ramp for Sea Surface Temperature Anomaly. Title and key scaffolding (TS) removed abbreviations and jargon from the titles, moved the key to the left side of the image to represent given information (Graham, 2002), and added United States customary units of measurement alongside metric units. The fifth version was fully-scaffolded (FS) and incorporated GS, CS and TS changes. See Appendix Figure 1 for an example of the unscaffolded and fully-scaffolded visualizations, and Table 1 for scaffolding types and versions of the visualizations.

Table 1
Scaffolding types per stimulus visualization version.

Stimulus Version	Month of Data Depicted	Number of Scaffolds	Types of Scaffolding
Unscaffolded (US)	January	None	None
Geographic Scaffolding (GS)	April	Single	Continent and ocean basin names added
Culturally-relevant Colors Scaffolding (CS)	April	Single	Continuous or divergent color scales as appropriate to the topic with hues chosen to match expectations.[a]
Title and Measurement Unit Scaffolding (TS)	July	Single	Removed abbreviations and jargon; added information about time span (one month). Added customary measurement units.
All Three	July	Fully-scaffolded (FS)	GS, CS, and TS

Note. [a] *See Appendix Figure 1 for scaffolded stimuli for each topic.*

Subjects were first calibrated to the SMI-RED™ eye-tracking system using the standard five-point calibration with four-point validation procedure. The background of the calibration was white and the target black, consistent with the white background of the stimulus visualizations. If necessary, calibration was repeated until average deviation was less than one degree of visual angle in each direction, and most subjects were calibrated to less than one-half degree.

The experimental environment is the same for all subjects. Subjects were seated in front of a 22" monitor and directed to sit as still as possible; no chin rest was used (Duchowski, 2007). The researcher monitored the subjects' movement in three dimensions with SMI ExperimentCenter™ RED Tracking Monitor software to ensure they remained within the eye-tracker capture area of 60 – 80 cm from the monitor (SensoMotoric Instruments, 2012). Data was collected at 120Hz.

Each subject was presented the five versions of the visualization for the topic that they were not shown in the clinical interview. Ultimately, based on what subjects were shown in the clinical interview, 3 expert and 2 novice subjects were shown the Sea Surface Temperature stimuli, 2 experts and 6 novices saw the Sea Surface Temperature Anomaly stimuli, and 3 experts and 2 novices saw the Chlorophyll Concentration stimuli.

Visualization versions were randomized for presentation using SMI Experiment Center™. Subjects were instructed that for each visualization, they would first have 10 seconds to look at the visualization without any direction, a period of "spontaneous looking" (SL) as described by Libarkin, Clark, and Simmon (2010). After the 10 seconds, the interview questions began. Visualizations were then presented to the subject as long as they were answering the interview questions, with the interviewer then advancing the stimulus manually. In between stimuli, subjects were presented with a noise image matched to the brightness and contrast of the visualizations.

Interview questions were an abbreviated version of the clinical interview in Stofer (2013). They asked the subjects to describe 1) the main idea of the visualization, 2) the meaning of the colors, 3) the measurement unit used, 4) the time span depicted, and 5) the season depicted in the visualization. After each answer, the subject was asked "How do you know" as an abbreviated probing session to reveal subject meaning-making.

Analysis was both quantitative and qualitative. SL periods were followed by the concurrent interview questions, which always included the initial question about the main idea (MI) of the stimulus. Length of total expo-

DOI 10.16910/jemr.7.5.2

ISSN 1995-8692

sure to the visualization varied due to the varying length of subject interview answers to these and subsequent questions. Therefore, only the SL and MI periods are analyzed here.

Quantitative Analysis

The raw data, output via SMI BeGaze™ software, consist of the subjects' gaze coordinates, pupil moving direction, and duration of the gaze. Three types of responses are extracted from this raw data: fixation, which measures the coordinated points at which subjects gazed longer than 80 ms; duration, the length of time fixation occurred, and scan paths, or the trace in which the eyes travels between fixations. We chose 80 ms as the temporal fixation threshold as a compromise between the text at 60 ms (Rayner, 1998), and image at 100 ms (Manor & Gordon, 2003), portions of the visualization. Spatial dispersion was 100 pixels maximum (Johansson, Holsanova, & Holmqvist, 2011). We only report left eye data.

For the visualizations, three areas of interest (AOI) were drawn: the map portion with data overlay, the title, and the key to the map including color bar and measurement unit labels. Two other AOI were drawn to verify the differences in map size and placement of the key between the "larger" (US, CS, and GS versions) and "smaller" maps (TS, FS) were irrelevant. On the larger maps, an overlap AOI covering the left part of the map where the key was placed in the smaller maps, and in the smaller maps, a color cutout polygon was drawn to exclude portions of the map that were encompassed by the key AOI. As there were few fixations falling in the left-hand portion of the map based on these two AOI, they were excluded from further analysis and the standard rectangular AOI were considered acceptable despite the map differences. In addition, the AOI calculations included fixations that were not in the AOI themselves, namely fixations in White Space, that is, on-screen but outside the three defined areas, and Off-screen. As these two categories also represented minimal numbers of fixations, their analysis was discarded.

Our main independent variable of interest was expertise as expert professional scientist versus novice non-professional adult. Three additional independent variables were investigated as potential variables of interest based on experimental design: topic, SST, SST Anomaly, or chlorophyll; scaffolding level, US, CS, GS, TS, FS; and trial number to check for learning effects. We examined the data using graphical tools, and checked modeling assumptions of normality, collinearity, autocorrelation and homoscedasticity. We used truncated linear regression to model the dependent data in spontaneous looking stages. Multinomial logistic regression was conducted to evaluate relative probabilities of fixation on given AOI using the above mentioned four independent variables.

In the eye-tracking SL stage, the distribution for the durations of fixation for all subjects had a median of 259 ms and was highly right-skewed. The right tail extended beyond 2000 ms, but the left tail was cut at 80 ms, as restricted by the definition of fixation. To counterbalance this skewness, we modeled the durations using a truncated regression model (Amemiya, 1973; Hausman & Wise, 1978). The truncated regression assumed there were fixations with durations less than 80 ms, and the full data followed a normal distribution. Since part of the duration data was systematically missing because of their values, the full data with all dependent and independent values were unobserved. We chose the truncated regression model over other types of analytical tools for skewed data, because it did not assume the underlying distribution has the same mean and variance, as required by Poisson regression, and also did not assume both tail densities converge to zero, as assumed by negative binomial regression.

Since there were always more than two AOI per visualizations, we assumed the percentage of fixations that fell into a certain AOI in a visualization followed a multinomial distribution, with the summation of percentages of all AOI equal to 1. We used the multinomial logistic regression model to calculate the effect of each independent variable on the variability (risk) for each AOI. The relative risk ratios (RRR) were reported along with its standard deviation and p-values. A RRR of, for instance, scaffolding over the AOI "Title", can be calculated as RRR = [A/(A+B)]/[C/(C+D)] = E, with variables A-D as shown in Table 2.

Table 2
Relative Risk Ratio calculation.

Risk	Title	NOT Title
Fully-scaffolded	A	B
Unscaffolded	C	D

Note. "Risk" in this case means chance of looking at a particular AOI in the denoted stimulus version.

DOI 10.16910/jemr.7.5.2

ISSN 1995-8692

The relative risk ratio means that, if given all other independent variables fixed, it is E-times more likely for someone to have a fixation on the "Title" AOI for a fully-scaffolded visualization than on the "Title" AOI for an unscaffolded one.

In the MI stage, we studied how the subjects varied in terms of their eye movement when asked to answer

"what is the main idea of this visualization?" concurrently with eye-tracking. All the subjects were pooled across topics to be studied together, and then novice and expert groups were also investigated separately. For fixations, the same four independent variables as the SL study, expertise, topic, scaffolding level, and trial, were included in the models. We also conducted the multinomial regression for the AOI distributions for these four variables in the MI stage, comparing the unscaffolded visualizations against single scaffoldings (color, geographic labels, title and key), and unscaffolded against fully scaffolded visualizations.

We checked the data against the assumptions of linear models and all assumptions were met. Durations of fixations were approximately normally distributed with a truncation from the left side. We believed that our data were collected so that each subject and test trial is independent from the others. No collinearity among explanatory variables was detected, and the variance of the data exhibited homoscedasticity. Similar to the findings of other papers (Tatler & Vincent, 2008), we discovered the fixations exhibit a moderate degree of autocorrelation within each trial for both SL and MI stages. This means the duration for one fixation is associated with durations for subsequent ones. Usually this suggests a time series component in the model. However, since we were interested in neither time as a variable of interest nor the progression or changes of durations for a given trial, we did not include time in our model and will study the time series effect of durations in a future work.

Qualitative Analysis

Qualitative analysis started with production of scan path images and heat maps for the SL condition only using SMI BeGaze™ analysis software. Due to an unequal number of subjects viewing each topic and the small number of subjects in general, to create these qualitative analysis products, data from all topics was overlaid together onto the versions of Sea Surface Temperature visualizations. As the placement of individual features of the overall visualizations was constant across topics, including the map itself, the title, the key, and the general geography, we looked at patterns of eye-tracking on these larger features. We could not, however, examine patterns within the ocean data, where the variations in data patterns in the ocean across seasons within the same topic and across topics were presented.

Experts and novices were compared on these qualitative images for the unscaffolded and fully-scaffolded cases. As expert gaze patterns did not seem to differ between these cases, nor were they expected to, we did not examine their unscaffolded and singly-scaffolded cases. However, novice gaze patterns for each type of scaffolding were compared to the unscaffolded case to judge improvement in meaning-making based on the individual scaffolds.

Results

Results showed overall differences in fixation duration, placement, and order on these visualizations in most cases, between experts and novices, evidenced in both quantitative and qualitative results.

Quantitative

SL Stage. We report here first the results from the spontaneous looking (SL) stage, followed by the main idea (MI) stage. Within each stage, duration of fixations was modeled in truncated regression models, and numbers of fixations within different AOI were fitted over multinomial distribution, as described in the Methods Section. Mean and median fixation duration in the SL stage was similar for both experts (M = 329 ms, median = 259) and novices (M = 326 ms, median 259 ms).

Among all subjects in the SL stage, only trial was a significant explanatory variable for the durations of fixations observed ($p < 0.001$). Expertise, level, and topic were not significant after accounting for the variations of the other independent variables. When the effect of trials was considered linearly, each additional trial on the visualization will reduce the mean duration of fixations by 16.2 ms (SD 1.71 ms). The effect size of the test, which is measured by Cohen's f2 statistic, is 0.00341, giving a statistical power of 0.79, a large effect (Cohen, 2013).

Across all stimuli in the SL stage, there were 2261 total fixations. The majority of fixations fell within the vis-

DOI 10.16910/jemr.7.5.2

ISSN 1995-8692

ualization areas (1236 for larger map, 740 for smaller map), and approximately 10% of fixations fell outside of any AOI, onto white space but still on screen (226 of 2261).

We hypothesized that the fixation placement over AOI is associated with the independent variables of not only topic but also expertise and level of scaffolding. We examined three AOI: Title, Key, and White Space.

The numbers of fixations for each AOI were compared in logistic regression between unscaffolded (US) and geographic scaffolding (GS), US and color scaffolding (CS), US and title/ key scaffolding (TS), and finally, US and full scaffolding (FS). Expertise did not explain much difference. For Level and Topic as effects, the US-GS and US-CS comparison did not yield a significant independent variable, but for the US-TS comparison, the different scaffolding levels were significant in explaining the shift of fixations in Title, Key, and White Space AOI (all $p < .001$). See Table 3.

Table 3
Relative Risk Ratios for Areas of Interest, Unscaffolded versus Title and Full Scaffolding, All Subjects (n = 16), SL.

	Title Scaffolding				Full Scaffolding			
	Level		Topic		Level		Topic	
AOI	Coefficient (*SE*)	RRR (95% CI)	Coefficient (*SE*)	RRR (95% CI)	Coefficient (*SE*)	RRR (95% CI)	Coefficient (*SE*)	RRR (95% CI)
Title	0.49 (0.9) ***	1.63 (1.36, 1.95)	-0.37 (0.18)*	0.69 (0.48, 0.99)	.23 (.06) ***	1.26 (1.12, 1.41)	-0.77 (0.21) ***	0.46 (0.31,0.7)
Key	.22 (.07) **	1.25 (1.09, 1.44)	-0.46 (0.16)**	0.63 (0.46, 0.86)	.16 (.04) ***	1.17 (1.1, 1.28)	-0.61 (0.17)***	0.54 (0.39,0.75)
White Space	.45 (.08) ***	1.57 (1.34, 1.83)	-0.2 (0.16)	(0.82) (0.6, 1.12)	.28 (.05) ***	1.32 (1.2, 1.45)	0.03 (0.15)	1.03 (0.76, 1.39)

*Note. SE = Standard Error, CI = Confidence Interval *p < .05. **p < .01. *** p < .001.*

The US-FS comparison with Level and Topic revealed the most evidence of association between distribution of fixations and our independent variables. Numbers of fixations, across all AOI, were highly associated with US-FS scaffolding difference (all $p < .001$ except White Space for Topic). See Table 4.

Table 4
Relative Risk Ratios for Areas of Interest with Expertise as Main Effect, Unscaffolded versus Scaffolded Cases, All Subjects (N = 16), MI.

	Geographic Scaffolding		Color Scaffolding		Title Scaffolding		Full Scaffolding	
AOI	Coefficient (*SE*)	RRR (95% CI)	Coefficient (*SE*)	RRR (95% CI)	Coefficient (*SE*)	RRR (95% CI)	Coefficient (*SE*)	RRR (95% CI)
Title	-1.31 (0.346) ***	0.27 (0.14,0.53)	-0.89 (0.285) **	0.41 (0.24,0.72)	-1.2 (0.237) ***	0.3 (0.19,0.48)	-0.3 (0.193)	0.74 (0.51,1.08)
Key	-0.13 (0.202)	0.88 (0.59,1.3)	-0.06 (0.183)	1.06 (0.74,1.51)	0.07 (0.22)	1.07 (0.7,1.65)	-0.63 (0.207) **	0.53 (0.36,0.79)
White Space	-0.39 (0.249)	0.68 (0.41,1.1)	0.22 (0.231)	1.24 (0.79,1.95)	-0.86 (0. 2) ***	0.42 (0.29,0.63)	-0.49 (0.182) **	. 0.61 (0.43,0.88)

*Note. SE = Standard Error , RRR = Relative Risk Ratio, CI = Confidence Interval. **p < .01. *** p < .005.*

Additionally, the model showed expertise was significantly associated with fixation change on Key AOI ($p < .05$).

MI Stage. There were a total of 3313 fixations in the Main Idea (MI) stage of the study. The mean for all participants was 41 fixations per trial with a median of 33

DOI 10.16910/jemr.7.5.2

ISSN 1995-8692

fixations. For novice subjects, the mean fixation per trial was 37 (median 30). For the expert group, the mean was 47 fixations per trial (median 44). The number of fixations was not significantly different between the two groups (*Ws* (n_1 = 35, n_2 = 45) = 942.5, p = .113). Average fixation duration for MI did vary significantly between the expert (M = 343 ms) and the novices (M = 401 ms), *Ws* (n_1 = 1653, n_2 = 1660) = 815, p = .008.

For the truncated regression of the durations of fixations among all subjects, expertise ($p < .001$) turned out to be statistically significant. The expert category was associated with a shorter duration of mean fixations (coefficient of estimation = 58.3 ms, SD = 13.2 ms, t = -4.42). Due to the increased numbers of fixations, the power of the test improved to 1.0, with a Cohen's f2 statistic of 0.009.

AOI analysis in the MI stage yielded expertise as the single most important variable explaining the shift of fixations among AOI in both US-FS and US-CS/US-GS/US-TS scaffolding comparisons. In US-FS comparison, expertise was significant in determining the RRR of Key and White Space AOI ($p < .01$). In all three US-Single scaffolding comparisons (US-GS, US-CS and US-TS), expertise was significant in determining the RRR of Title AOI ($p < .01$) and in determining the shift in White Space AOI in the US-TS comparison ($p < .005$). See Table 4. The increased chance of looking at white space may reflect the conservative size of the visualization element AOI; subjects may have been using peripheral vision rather than fixating directly on the visualization element (Kim, Dong, Xian, Upatising, & Yi, 2012).

In summary, the quantitative differences and similarities in number and duration of fixations among subject groups (by expertise), scaffolding level, trial, and topic of the visualizations complement and extend the interview data from Stofer (2013). Specifically, they show what differences exist in patterns of attention as measured by the time spent and object of gaze when looking at the various versions of the stimuli. Compared to novices, the experts generally spend less time per fixation and have more fixations per visualization when specifically asked to answer questions about the visualizations, indicating greater meaning-making (Holmqvist et al., 2011), in line with interview results (Stofer, 2013). Novices did change their patterns of looking in some cases with increased scaffolding or trials.

Qualitative

Heat map images clarified differences in expert and novice viewing of global data visualizations that were not always evident in the quantitative data. Heat maps visualize average duration of fixations on particular parts of stimuli using a color ramp to indicate duration. While in the SL case, there were no significant differences in dwell time at any scaffolding level between experts and novices, qualitative analysis did show qualitative differences in the focal areas of each group. On the unscaffolded images, novices spent more time on the Western Hemisphere of the map, specifically the northern part of the Western Hemisphere, than did experts, who tended to look at the entire image. See Appendix Figure 2.

However, with scaffolding, we noticed that novices scanned more of the map, showing heat maps that began to better resemble those of the experts, rather than concentrating so obviously on the quadrant containing the United States. This was particularly evident not only in the stimulus with geographic labels scaffolding but also in the fully-scaffolded version, which included geographic labels. In addition, in the fully-scaffolded case, the novices showed smaller fixations on the title than in the unscaffolded case, indicative of better meaning-making.

Resolution of the heat maps and areas of interest did not allow us to probe statistically whether the increased use of the map by novices in those scaffolding cases was due to increased use of the ocean data or simply examining the land geography more fully.

Heat map analysis of the novices' color-scaffolded case compared to the unscaffolded case paths was less informative as all three topics had different data patterns within the maps. However, the paths did indicate a change in viewing pattern with the color scaffolding. Whether this is in the direction of better meaning-making cannot be determined from this small sample. Heat map analysis of novices' title-scaffolded case compared to the unscaffolded case also indicated more time spent on the title and key. Each of these deserves further study.

Scan paths in turn confirmed more subtle differences in both expert and novice meaning-making with scaffolding. Scan paths visualize fixations in order, with a circle on the fixation point and lines tracking between fixations. Diameter of the circle represents dwell time. In particular, comparing the novices' unscaffolded and geographic scaffolded case shows that novices did use the geographic

DOI 10.16910/jemr.7.5.2

ISSN 1995-8692

labels to orient themselves. Also, comparing the experts' scan paths on the unscaffolded and title scaffolded cases showed they tended to focus on the first part of the title. This part of the title was consistently indicative of the time span represented; the question of understanding time span from the visualization was the one that most experts struggled with in the previous clinical interview experiment (Stofer, 2013).

Discussion

Our eye-tracking results show several details of the differences between expert and novice readers of global ocean data visualizations that were not evident from interviews alone. These data serve to confirm concurrent and previous clinical interview data, but these results also help to explain some of the remaining mysteries of the discrepancies in meaning-making between the groups.

First, we confirmed interview results that scaffolding improved novice meaning-making. While the number and duration of fixations did not show an effect of scaffolding level in spontaneous looking trials, AOI data showed increased probabilities of looking at the title and key in the cases (TS and FS) when those were scaffolded, and no change in the cases (CS and GS) where those elements were not scaffolded. When asked about the main idea, similar patterns in AOI emerged, though the singly-scaffolded cases were also significantly different. The most significant differences were in the case of the fully-scaffolded stimuli as expected, suggesting that more understandable elements overall support meaning-making using all the available elements, as subjects can then ably use multiple sources of information to make meaning.

The quantitative data did reveal expert-novice differences in the main idea stage. The difference of 58 ms represented 15% of the average fixation duration for novices, suggesting potential for improvement in dwell time associated with better meaning-making.

We also confirmed with qualitative scan path data that experts and novices were looking at the visualizations differently. We found limited use of the title and key by novices in the SL case, in line with a pilot study on a similar rainbow-hued global visualization (Libarkin et al., 2010). However, that study concluded novices' SL scan paths covered more of the visualizations than the experts' scan paths. This finding was opposite to the conclusions here for the unscaffolded scan paths, but the visualizations used by Libarkin et al. (2010) showed data both on land and in the ocean, and in that pilot study, neither experts or novices focused much at all on any of the ocean data. Since our visualizations were exclusively ocean data, and we were not told what disciplinary expertise the experts in Libarkin et al. (2010) had, those differences may be responsible for the discrepancies between studies. Our quantitative fixation data showing shorter fixations by experts on the main idea meaning task are in line with differences in these groups from other expert-novice eye-tracking comparisons on map reading (Çöltekin et al., 2010, 2009; Ooms et al., 2012), and on expert-novice differences in map reading in behavioral tasks (Allen, Miller Cowan, & Power, 2006; Anderson & Leinhardt, 2002; Gilhooly, Wood, Kinnear, & Green, 1988; Kastens et al., 2009).

Qualitative scan paths also changed for the novices in our study for all the types of scaffolding; with each type of scaffolding, the novices' paths started to resemble more closely the paths of the experts than in the unscaffolded case. In particular, the novices used more of the map overall when the geographic labels were added, suggesting they better oriented to the visualization and were instead able to focus on the meaning of the data and overall visualization. Focus maps showed novices also making use of the geographic labels in particular. This can be compared with the presence of task-irrelevant information proving a distraction in map reading (Canham & Hegarty, 2010); providing information in the form of labels allowed subjects to focus on the larger task. Focus maps showed experts, too, made use of the title to answer a question that many of them struggled to answer when the titles were not scaffolded. These findings also support the previous interview data (Stofer, 2013).

A new finding from the eye-tracking data was the effectiveness of trial on both groups. This *practice effect* was not as clearly evident in the clinical interview data, possibly due to the unequal number of trials subjects underwent in that experiment; due to both time constraints and recognition of the visualizations as the same data, not all subjects saw all ten stimuli in the clinical interviews (See Stofer (2013) for full details). The practice effect does align with clinical interview findings that novices were generally unfamiliar with these visualizations and the interpretation task and suggest that the enculturation of the experts and their training with similar tasks was

DOI 10.16910/jemr.7.5.2

ISSN 1995-8692

responsible for at least part of their superior performance. This effect was also concurrent with data on the importance of training to expertise (Benner, 1982; Chase & Simon, 1973; Ericsson et al., 1993).

In addition, when prompted to look at the visualization specifically to determine the main idea, fixation dwell times did significantly differ between the expert and novice groups, whereas they did not differ in the spontaneous looking condition. This suggested that the questions themselves in the interviews could actually shape novices' methods of investigating the data. A difference in probability of looking at the Title AOI based on the trial number could indicate practice effect or a comparison with previous visualizations to understand differences, both of which warrant further exploration.

Overall, our hypotheses regarding the effectiveness of scaffolding were confirmed, though the influence of some of the independent variables was revealed through different analyses than expected. These results were in line with conclusions from the interview portion of the experiment as reported in Stofer (2013), namely, that experts and novices made different meaning from the visualizations and novices in particular struggled with certain aspects of the visualizations and the tasks that experts have been trained to understand.

Conclusions

Taken together, the quantitative fixation data, area of interest analysis, and scan path analysis showed that experts and novices had different unconscious approaches to viewing the stimulus visualizations in both the unprompted, spontaneous looking stage and the question-prompted, main idea focused stage. While scaffolding in spontaneous looking stage did not change fixation dwell times, it did in fact change novice patterns of looking, adding to the lines of evidence suggesting the effectiveness of the scaffolding. The qualitative data also revealed information about how the interventions were working to improve novices' meaning-making. Altogether, eye-tracking provided not only triangulation of interview data, but also complementary findings that were not apparent in the interview alone.

Our expert-novice comparison revealed differences in cognitive processing by the two groups when examining a global data visualization. The results, in accordance with similar findings on related but distinct map-reading tasks warrant further investigation into the particular differences. Specifically, the way the two groups look at the data patterns themselves should be explored. Finally, the results from the two groups suggest ways we can improve the visualizations for novice users through scaffolding.

Combined with interview data from Stofer (2013), we recommend specific internal scaffolds for these visualizations to support meaning-making by broader audiences. In particular, those scaffolds should include those tested here, namely geographic labels, audience-appropriate color schemes, and jargon-free titles and measurement units. Additional internal scaffolds and external interventions may be warranted to allow further improvement. Future investigations could examine eye-tracking on patterns within the map portions of the visualizations themselves to investigate learning about the data depicted and supporting skills of data interpretation such as those in the Next Generation Science Standards (NGSS Lead States, 2013).

Acknowledgements

The authors wish to thank Shawn Rowe, Anthony Hornof, and Christy Steffke for assistance with and consultation on this research, two anonymous reviewers for feedback on this manuscript, and SMI for technical support. This work was supported by the National Science Foundation grant 1114741 to Shawn Rowe and a Curtis and Isabella Holt Marine Education Award to Kathryn Stofer.

References

Allen, G. L., Miller Cowan, C. R., & Power, H. (2006). Acquiring information from simple weather maps: Influences of domain-specific knowledge and general visual–spatial abilities. *Learning and Individual Differences*, *16*(4), 337–349. doi:10.1016/j.lindif.2007.01.003

Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica*, *41*(6), 997. doi:10.2307/1914031

Anderson, K. C., & Leinhardt, G. (2002). Maps as Representations: Expert Novice Comparison of Projec-

DOI 10.16910/jemr.7.5.2

ISSN 1995-8692

tion Understanding. *Cognition & Instruction*, *20*(3), 283–321. doi:Article

Benner, P. (1982). From novice to expert. *The American Journal of Nursing*, *82*(3), 402–407.

Bransford, J., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How People Learn: Brain, Mind, Experience, and School* (Expanded edition.). Washington, DC: National Academies Press.

Bruer, J. T. (2006). On the implications of neuroscience research for science teaching and learning: Are there any? A skeptical theme and variations: The primacy of psychology in the science of learning. *CBE Life Sciences Education*, *5*(2), 104–110.

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, *20*(2), 155–166. doi:10.1016/j.learninstruc.2009.02.014

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–61.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices*. *Cognitive Science*, *5*(2), 121–152. doi:10.1207/s15516709cog0502_2

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. Hoboken, NJ: Routledge Academic.

Çöltekin, A., Fabrikant, S. I., & Lacayo, M. (2010). Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings. *International Journal of Geographical Information Science*, *24*(10), 1559–1575. doi:10.1080/13658816.2010.511718

Çöltekin, A., Heil, B., Garlandini, S., & Fabrikant, S. I. (2009). Evaluating the Effectiveness of Interactive Map Interface Designs: A Case Study Integrating Usability Metrics with Eye-Movement Analysis. *Cartography and Geographic Information Science*, *36*(1), 5–17. doi:10.1559/152304009787340197

Daniel, D. B. (2012). Promising principles: Translating the science of learning to educational practice. *Journal of Applied Research in Memory and Cognition*, *1*(4), 251–253. doi:10.1016/j.jarmac.2012.10.004

Duchowski, A. T. (2007). *Eye Tracking Methodology* (Second.). London: Springer Science+Business Media.

Eberbach, C., & Crowley, K. (2009). From everyday to scientific observation: How children learn to observe the biologist's world. *Review of Educational Research*, *79*(1), 39 –68. doi:10.3102/0034654308325899

Edelson, D. C., & Gordin, D. (1997). Creating science learning tools from experts' investigation tools: A design framework. Presented at the Annual Meeting of the National Association for Research in Science Teaching, Oakbrook, IL.

Edelson, D. C., & Gordin, D. (1998). Visualization for learners: A framework for adapting scientists' tools. *Computers & Geosciences*, *24*(7), 607–616.

Elby, A. (2001). Helping physics students learn how to learn. *Physics Education Research, American Journal of Physics*, *69*(S1), S54–S64.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406. doi:10.1037/0033-295X.100.3.363

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: evidence of maximal adaption to task constraints. *Annual Review of Psychology*, *47*, 273+.

Ertmer, P. A., & Newby, T. J. (1996). The expert learner: Strategic, self-regulated, and reflective. *Instructional Science*, *24*(1), 1–24.

Filippini-Fantoni, S., Jaebker, K., Bauer, D., & Stofer, K. (2013). Capturing visitors' gazes: Three eye tracking studies in museums. In N. Proctor & R. Cherry (Eds.), *Museums and the Web 2013*. Silver Spring, MD: Museums and the Web. Retrieved from http://mw2013.museumsandtheweb.com/paper/capturing-visitors-gazes-three-eye-tracking-studies-in-museums/

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 431–446. doi:10.1037/0096-1523.28.2.431

Gegenfurtner, A., Siewiorek, A., Lehtinen, E., & Säljö, R. (2013). Assessing the Quality of Expertise Differences in the Comprehension of Medical Visualizations. *Vocations and Learning*, *6*(1), 37–54. doi:10.1007/s12186-012-9088-7

DOI 10.16910/jemr.7.5.2

ISSN 1995-8692

Gilhooly, K. J., Wood, M., Kinnear, P. R., & Green, C. (1988). Skill in map reading and memory for maps. *The Quarterly Journal of Experimental Psychology*, *40A*, 87–107.

Graham, L. (2002). *Basics of Design: Layout and Typography for Beginners*. Albany, NY: Delmar.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science*, *14*(5), 462–466. doi:10.2307/40064168

Haley Goldman, K., Kessler, C., & Danter, E. (2010, September). Science on a Sphere: Cross-site summative evaluation. Institute for Learning Innovation. Retrieved from http://www.oesd.noaa.gov/network/SOS_evals/SOS_Final_Summative_Report.pdf

Hausman, J. A., & Wise, D. A. (1978). A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica*, *46*(2), 403–426.

Hegarty, M. (2013). Cognition, metacognition, and the design of maps. *Current Directions in Psychological Science*, *22*(1), 3–9. doi:10.1177/0963721412469395

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye Movements: A Window on Mind and Brain*. Oxford, UK: Elsevier Ltd.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Weijer, J. van de. (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press.

Jee, B. D., Uttal, D. H., Spiegel, A., & Diamond, J. (2013). Expert–novice differences in mental models of viruses, vaccines, and the causes of infectious disease. *Public Understanding of Science*, 0963662513496954.

Jensen, E. (1998). *Teaching with the brain in mind*. Alexandria, VA: Association for Supervision and Curriculum Development.

Johansson, R., Holsanova, J., & Holmqvist, K. (2011). The dispersion of eye movements during visual imagery is related to individual differences in spatial imagery ability. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 1200–1205). Austin, TX.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect. *Educational Psychologist*, *38*(1), 23–31. doi:10.1207/S15326985EP3801_4

Kastens, K. A., Agrawal, S., & Liben, L. S. (2009). How Students and Field Geologists Reason in Integrating Spatial Observations from Outcrops to Visualize a 3-D Geological Structure. *International Journal of Science Education*, *31*(3), 365 – 393.

Krygier, J., & Wood, D. (2011). *Making Maps: A Visual Guide to Map Design for GIS*. Guilford Press.

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and Novice Performance in Solving Physics Problems. *Science*, *208*(4450), 1335–1342. doi:10.2307/1684057

Libarkin, J. C., Clark, S. K., & Simmon, R. B. (2010, October 31). *The color of confusion in an expert world*. Presented at the Geolgocial Society of America, Denver, CO.

Libarkin, J. C., Thomas, S. R., & Ruetenik, G. (2013). Visual salience in climate change imagery is in the eye of the beholder. Presented at the Eye-tracking Mini-Conference, Lansing, MI. Retrieved from http://create4stem.msu.edu/sites/default/files/event/files/LibarkinThomasRuetenik4_13_13.pdf

Lobben, A. K., Lawrence, M., & Olson, J. M. (2009). fMRI and human subjects research in cartography. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *44*(3), 159–169.

Lobben, A. K., Olson, J. M., & Huang, J. (2005). Using fMRI in cartographic research. In *Proceedings of the 22nd International Cartographic Conference* (p. 10). A Coruna, Spain.

Manor, B. R., & Gordon, E. (2003). Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *Journal of Neuroscience Methods*, *128*(1–2), 85–93. doi:10.1016/S0165-0270(03)00151-1

Middendorf, J., & Pace, D. (2004). Decoding the disciplines: A model for helping students learn disciplinary

DOI 10.16910/jemr.7.5.2

ISSN 1995-8692

ways of thinking. *New Directions for Teaching and Learning*, *2004*(98), 1–12.

National Oceanic and Atmospheric Administration. (n.d.). Science on a Sphere. Retrieved October 23, 2012, from sos.noaa.gov

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press. Retrieved from http://www.nextgenscience.org/next-generation-science-standards

Ooms, K., De Maeyer, P., & Fack, V. (2014). Study of the attentive behavior of novice and expert map users using eye tracking. *Cartography and Geographic Information Science*, *41*(1), 37–54.

Ooms, K., De Maeyer, P., Fack, V., Van Assche, E., & Witlox, F. (2012). Interpreting maps through the eyes of expert and novice users. *International Journal of Geographical Information Science*, *26*(10), 1773–1788. doi:10.1080/13658816.2011.642801

Phipps, M., & Rowe, S. M. (2010). Seeing satellite data. *Public Understanding of Science*, *19*(3), 311 –321. doi:10.1177/0963662508098684

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372.

Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual span in expert chess players: Evidence from eye movements. *Psychological Science*, *12*, 49–56.

Rottman, B. M., Gentner, D., & Goldwater, M. B. (2012). Causal systems categories: Differences in novice and expert categorization of causal phenomena. *Cognitive Science*, *36*(5), 919–932.

Rowe, S. M., Stofer, K., Barthel, C., & Hunter, N. (2010). *Hatfield Marine Science Center Magic Planet Installation Evaluation Findings*. Corvallis, OR: Oregon Sea Grant.

Schubert, C. C., Denmark, T. K., Crandall, B., Grome, A., & Pappas, J. (2013). Characterizing novice-expert differences in macrocognition: an exploratory study of cognitive work in the emergency department. *Annals of Emergency Medicine*, *61*(1), 96–109.

SensoMotoric Instruments. (2012, March). ExperimentCenter Manual version 3.1. SensoMotoric Instruments.

Sheridan, H., & Reingold, E. M. (2014). Expert vs. novice differences in the detection of relevant information during a chess game: evidence from eye movements. *Frontiers in Psychology*, *5*. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4142462/

Steffke, C. L., & Libarkin, J. C. (2012, November 4). *Guiding symbology and display selection to produce more effective images for conveying information*. Poster presented at the Geological Society of America, Charlotte, NC.

Steffke, C. L., & Libarkin, J. C. (2013, October 28). *Which colors are better: An eye tracking study of color ramp symbology*. Paper presented at the Geological Society of America 2013 Annual Meeting, Denver, CO.

Steinke, T. R. (1987). Eye Movement Studies In Cartography And Related Fields. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *24*(2), 40–73. doi:10.3138/J166-635U-7R56-X2L1

Stofer, K. A. (2013, April 29). *Visualizers, Visualizations, and Visualizees: Differences in Meaning-Making by Scientific Experts and Novices from Global Visualizations of Ocean Data* (Doctoral Dissertation). Oregon State University, Corvallis, OR.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2*, 1–18.

Van Gog, T. (2006). *Uncovering the problem-solving process to design effective worked examples* (Doctoral Dissertation). OpenUniversiteitNederland, Heerlen, The Netherlands.

Walsh, C. M., Rose, D. N., Dubrowski, A., Ling, S. C., Grierson, L. E. M., Backstein, D., & Carnahan, H. (2011). Learning in the Simulated Setting: A Comparison of Expert-, Peer-, and Computer-Assisted Learning: *Academic Medicine*, *86*, S12–S16. doi:10.1097/ACM.0b013e31822a72c7

Wolf, S. F. (2012). Expert and novice categorization of introductory physics problems. Retrieved from http://adsabs.harvard.edu/abs/2012PhDT.....90W

DOI 10.16910/jemr.7.5.2

ISSN 1995-8692

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100. doi:10.1111/j.1469-7610.1976.tb00381.x

Appendix Figures

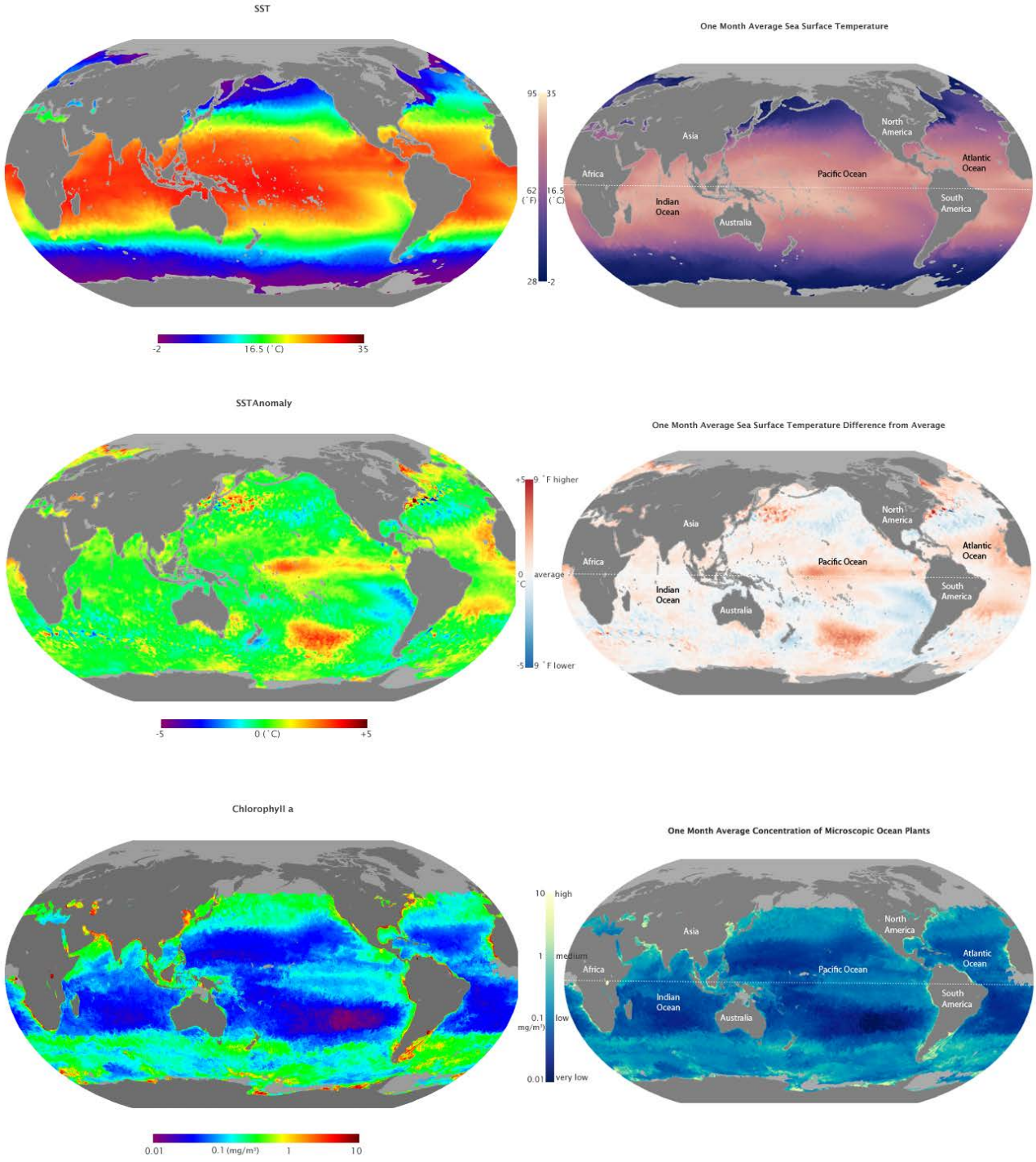
Appendix Figure 1. Appendix Figure 1. Unscaffolded and fully-scaffolded visualizations. Left (top to bottom): rainbow color scale versions of Sea Surface Temperature (SST), SST Anomaly, and Chlorophyll. SST and Chlorophyll each show a single continuous variable; SST Anomaly shows a diverging variable of higher-than-average and lower-than-average scales. Right: improved color schemes, titles and keys, and geographic labels for the same.

Appendix Figure 2. Heat maps showing all Novice (top left) and all Expert (bottom left) fixations in the first 10 seconds of viewing on unscaffolded visualizations (Spontaneous Looking). On right, novice (top) and expert (bottom) heat maps for scaffolded visualizations, SL

DOI 10.16910/jemr.7.5.2

ISSN 1995-8692

Appendix Figure 1. Unscaffolded and fully-scaffolded visualizations. Left (top to bottom): rainbow color scale versions of Sea Surface Temperature (SST), SST Anomaly, and Chlorophyll. SST and Chlorophyll each show a single continuous variable; SST Anomaly shows a diverging variable of higher-than-average and lower-than-average scales. Right: improved color schemes, titles and keys, and geographic labels for the same.



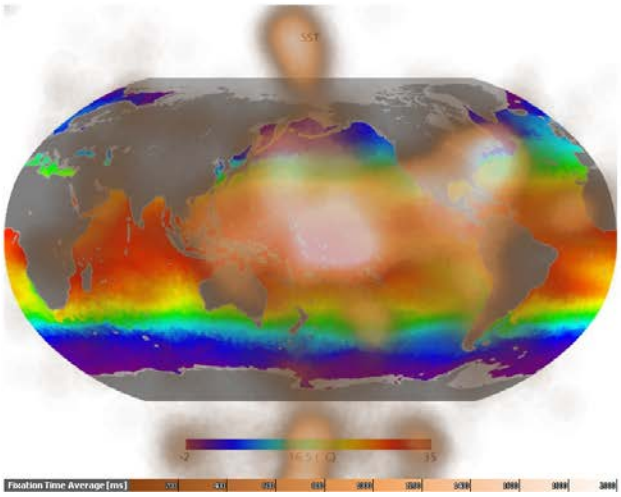
DOI 10.16910/jemr.7.5.2

ISSN 1995-8692

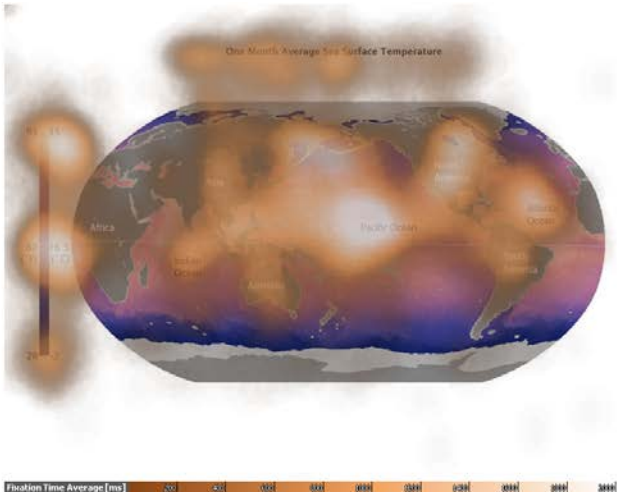
Appendix Figure 2. Heat maps showing all Novice (top left) and all Expert (bottom left) fixations in the first 10 seconds of viewing on unscaffolded visualizations (Spontaneous Looking). On right, novice (top) and expert (bottom) heat maps for scaffolded visualizations, SL.

All Novices, Spontaneous Looking

Unscaffolded

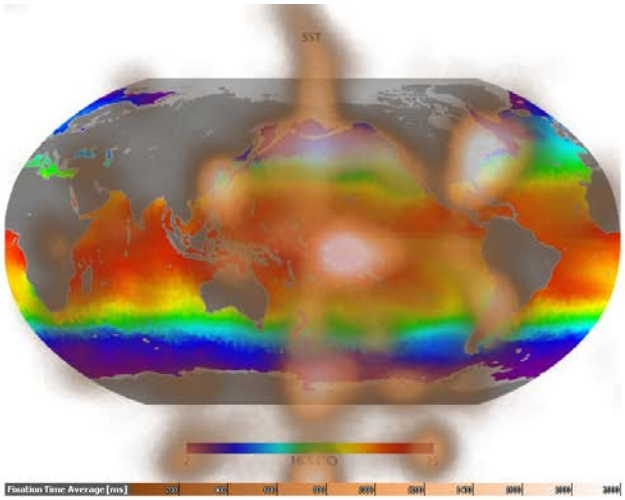


Fully-Scaffolded

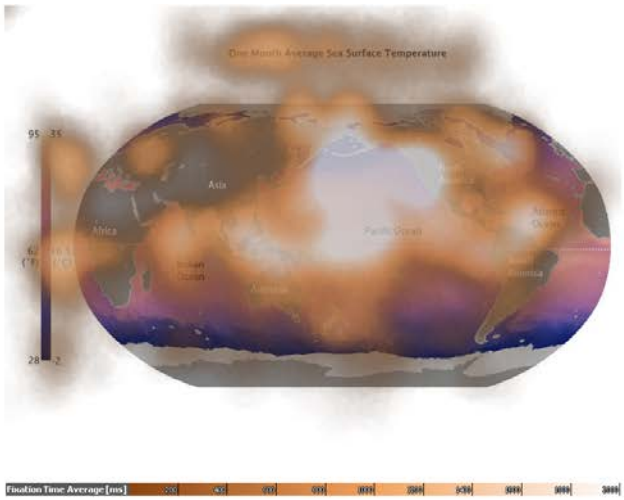


All Experts, Spontaneous Looking

Unscaffolded



Fully-Scaffolded



DOI 10.16910/jemr.7.5.2

ISSN 1995-8692