JOURNAL OF GEOSCIENCE EDUCATION **64**, 231–241 (2016)

When a Picture Isn't Worth 1000 Words: Learners Struggle to Find Meaning in Data Visualizations

Kathryn A. Stofer[1,a]

ABSTRACT

The oft-repeated phrase "a picture is worth a thousand words" supposes that an image can replace a profusion of words to more easily express complex ideas. For scientific visualizations that represent profusions of numerical data, however, an untranslated academic visualization suffers the same pitfalls untranslated jargon does. Previous research and commentary suggests scaffolding from cognitive, constructivist, and sociocultural traditions to promote meaning-making by broad audiences, but limited empirical research examines the effectiveness of these scaffolds for adult learners viewing global ocean data. Five versions of visualizations including translating titles and measurement units, altering color schemes, and geographic labels were presented on three topics to expert oceanographers and novice nonscientists. Qualitative analysis of semistructured clinical interviews suggest that these scaffolds assist these audiences but are not sufficient for novices to make meaning similar to experts without further instruction or assistance in interpreting and judging patterns of data in visualizations. © *2016 National Association of Geoscience Teachers*. [DOI: 10.5408/14-053.1]

Key words: data visualization, global satellite data, scaffolding, pattern recognition

INTRODUCTION

Visualizations of scientific data (see Fig. 1) represent numerical data in a form that ostensibly makes it easier to recognize patterns and make judgments on the data. The oft-repeated phrase "a picture is worth a thousand words" supposes that a communicator can use a visual representation in place of a profusion of words to more easily express complex ideas. However, when it comes to visualization of data, the academic scientist who wishes to communicate ideas may fall into the trap of thinking that just because the data is in a visual, image-based format, it is easy for everyone to understand, no matter their academic science background. An untranslated academic visualization used to communicate with other professionals often suffers the same sorts of pitfalls that scientific jargon does (Light and Bartlein, 2004; Phipps and Rowe, 2010). Visual representations of data can have embedded jargon, cultural context, and complexity in the same way an academic journal article can. This article investigates meaning-making by nonscience-major novice undergraduates from global ocean data visualizations with various types of cultural conventions and compares their meaning-making to that of experts in oceanography.

The Importance of Analyzing Data Visualizations

Increasingly, scientists wish to use data visualizations to convey information, especially through rapidly spreading technologies such as NOAA's Science on a Sphere™ (Haley Goldman et al., 2010) and classroom versions of spherical globes, Google Earth™, or even in documentary films. Educators in these settings, when they exist, may feel ill equipped to make use of untranslated visuals, lacking the time, resources, or permissions to alter visualizations to make them more in line with learners' current development and background (Barthel, 2010). Yet, there is a critical need to help students become adept at spatial thinking (Committee on the Support for Thinking Spatially, 2006; Kastens and Ishikawa, 2006; Goodchild and Janelle, 2010). Climate data, in particular, is often presented as global averaged data overlaid on a world map or data differences from average, and presumed to make the case to the general public that climate change is occurring. Visual presentations can indeed be compelling and help overcome alternative conceptions (Nyhan and Reifler, 2014), but the visuals must themselves be clear to the viewer.

School-age learners show improved meaning-making from visualizations with scaffolding (Wood et al., 1976), particularly for chemical molecules and associated models (e.g., Kozma et al., 2000; Chang and Linn, 2013), and with scaffolds primarily external to the visualizations themselves, in the form of additional explanatory text, structured problem solving steps, or instruction. Recent empirical work with visualizations at the intersection of cognitive sciences and geosciences (Kastens and Ishikawa, 2006; Fabrikant and Lobben, 2009; Fabrikant et al., 2010; Hegarty, 2011, 2013; Steffke and Libarkin, 2012, 2013) and specifically, global ocean satellite data visualizations (Phipps and Rowe, 2010; Rowe et al., 2010; Stofer and Che, 2014), have begun to examine scaffolds that are integral parts of the representation of the data itself or make changes to titles or legends. These "internal" scaffolds are intended to allow meaning-making when the visualizations are presented in a more stand-alone fashion, such as one might encounter in a media story or science center exhibit. Understanding visual representations is a learned skill (Cid et al., 2009), but many experts do not even encounter global satellite data visualizations until graduate study in particular fields such as oceanography (Stofer, 2013). However, data visualizations are becoming more pervasive in nonprofessionals' lives (Abelson, 2013; Tversky, 2014).

Several studies have also examined the use of some types of visualizations by audiences with little scientific background compared to experts. For example, students with more

Received 22 September 2014; revised 11 August 2015 and 18 December 2015; published online 19 August 2016.

[1]Department of Agricultural Education and Communication, University of Florida, P.O. Box 110540, Gainesville, Florida 32611, USA

[a]Author to whom correspondence should be addressed. Electronic mail: stofer@ufl.edu. Tel.: 352-273-3690. Fax: 352-392-9585.

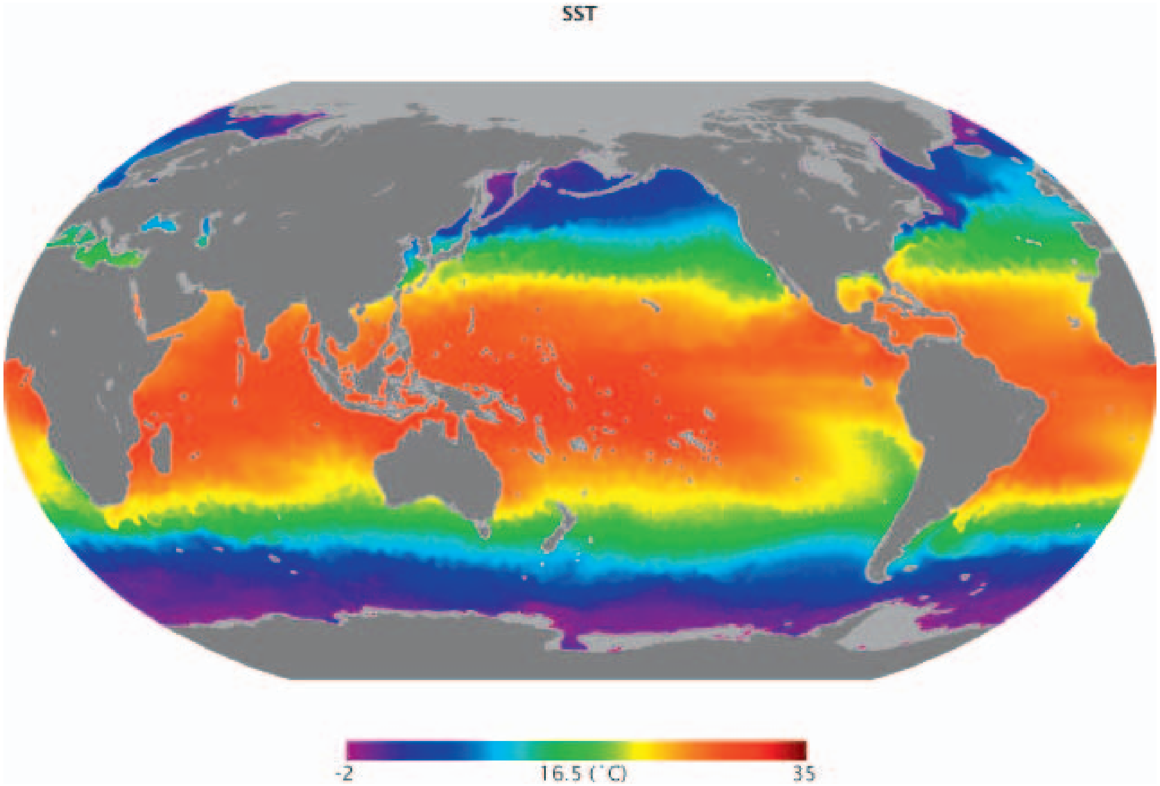

FIGURE 1: Sea surface temperature satellite data overlaid on a global map as typically presented to experts, without any culturally familiar supporting information added (unscaffolded). The color figure can be viewed in the online article.

background knowledge looked at different parts of a cellular transport diagram than students with less knowledge (Cook et al., 2008), novice students have expectations of color schemes that may or may not transfer to new representations (Cid et al., 2009), and experts and novices attend to different parts of visualizations (Breslow et al., 2009; Libarkin, J. C., personal communication November 4, 2012). Expert oceanographers bring different cultural expectations to the interpretation of visualizations than novices with only a high school science background; the experts need fewer explicit labels and can overcome characteristics of color schemes that interfere with human perception to make meaning. The novices get stuck trying to interpret these details and spend less time trying to make sense of the more important patterns in the data (Stofer and Che, 2014).

Visualization design combines both perceptual and cultural characteristics, and relies on the viewer to construct meaning, building on prior knowledge and experience. Therefore, the conceptual framework for the scaffold design here also drew on theories of neuroscience and physics of perception (Serway and Faughn, 2009; Light and Bartlein, 2004), constructivism (Driver, 1995), and sociocultural learning theory (Vygotsky, 1978; John-Steiner and Mahn, 1996). These are similar scaffolds to those suggested in previous frameworks for student learners in school (Edelson and Gordin, 1997). Specifically, color choices were made to eliminate perceptual confusion that may be produced when yellow-green hues represent middle, rather than extreme, values, as yellow-green is perceptually most salient to the human eye (Serway and Faughn, 2009; Light and Bartlein, 2004). Measurement units and titles had jargon removed to better fit prior knowledge of a public audience (Driver, 1995). Overall color schemes were chosen that were expected to fit with cultural expectations of a Western audience (John-Steiner and Mahn, 1996; Conroy, 1998; Light and Bartlein, 2004; Breslow et al., 2009; Phipps and Rowe, 2010). Namely, temperatures were represented in shades ranging from purple (cool) to pink/white (warm), chlorophyll was represented in shades of green to indicate plant matter, and temperature differences from average were represented in blue shades for lower than average and red shades for higher than average temperatures (see Figs. 1 and 2).

This study, therefore, sought to directly compare novice adults with expert professionals using global visualizations of ocean data similar to the types of visualizations encountered by the general public in out-of-school or media experiences. Specifically, this meant creating visualizations scaffolded internally in the absence of context aside from a brief title and color legend, and empirically examining presumably culturally relevant color scales for oceanographic data by offering participants multiple versions of visualizations with different levels of scaffolding. The current investigation also sought to capture participants' judgments of time periods and seasons represented and asked participants to offer evidence in the visualization to support their claims. The visualizations used here are those referred to as Level 5 by Taber et al. (2012): ostensibly "easy-to-use/universal display-image data" (p. 251).

METHODS

Context

The study took place at a large public university in the Pacific Northwest region of the United States. The interviews

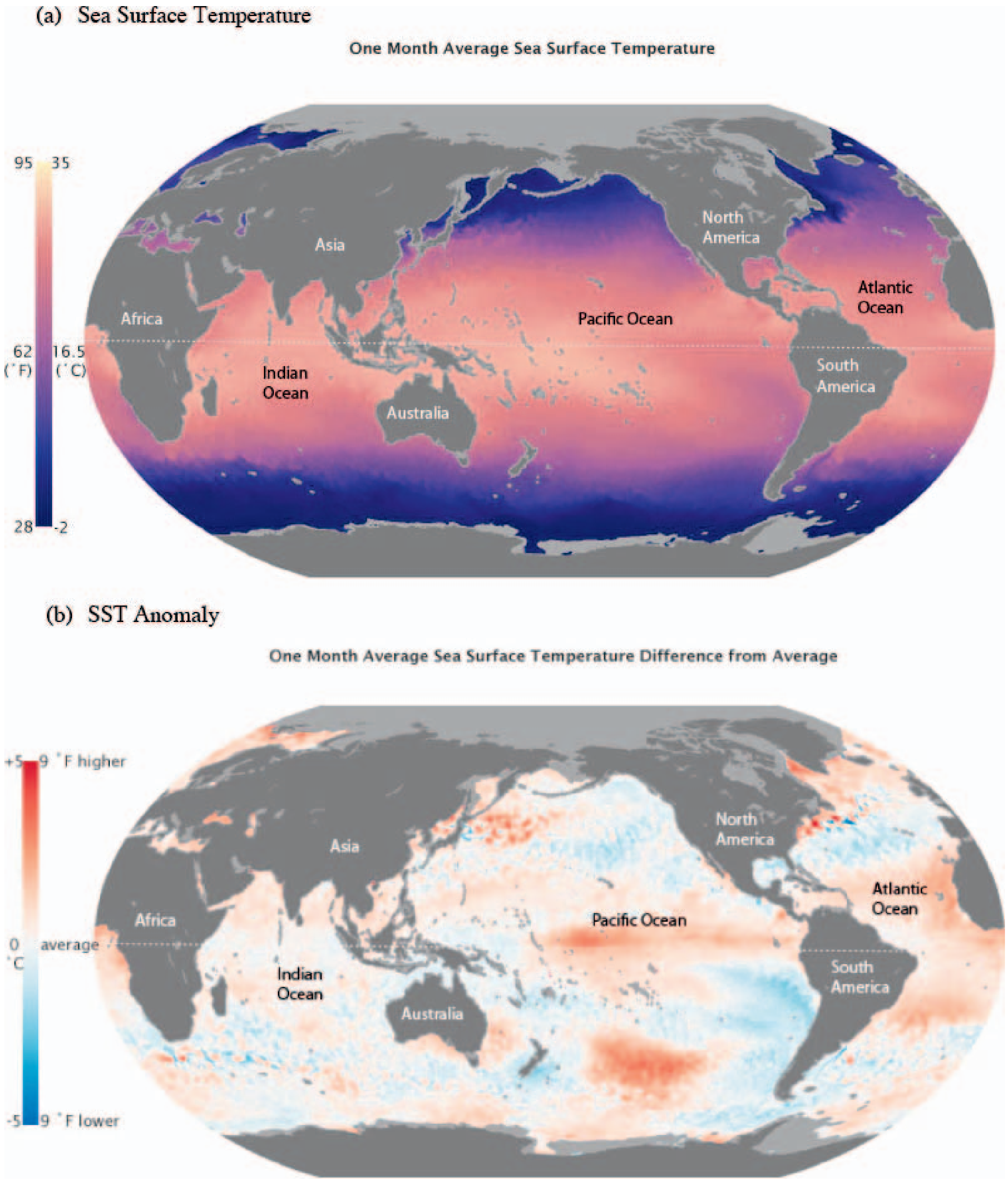

FIGURE 2: ***Full-Scaffolding*** **versions of the three topic visualizations presented to participants. Each fully scaffolded visualization includes** ***geography*****,** ***color*****, and** ***title and legend*** **scaffolding. (a) Sea Surface Temperature; (b) SST Anomaly, (c) Chlorophyll-a. The color figure can be viewed in the online article.**

described here were the first part of a larger study of meaning-making of visualizations; the entire study is described in Stofer (2013), and the second part of the study, involving eye tracking, is also detailed in Stofer and Che (2014). Participants were initially recruited for the interview portion, and then a subset of the group was invited back for the eye-tracking portion. For the entire study, the author used three topics of visualizations for stimuli: Sea Surface Temperature (SST), Sea Surface Temperature Anomaly (SST Anomaly), and Chlorophyll-a. Therefore, participants in the initial interviews described here were shown two of the three topic stimuli selected at random; the other topic for each participant was reserved in case they were to participate in the eye-tracking study.

Participants

The study participants were either expert oceanographers with at least five years of professional experience beyond the PhD or adult science novices, at least 18 years old, and with no more than two years of undergraduate study completed. Novices were not in science or engineering major programs. Expert participants were recruited using a

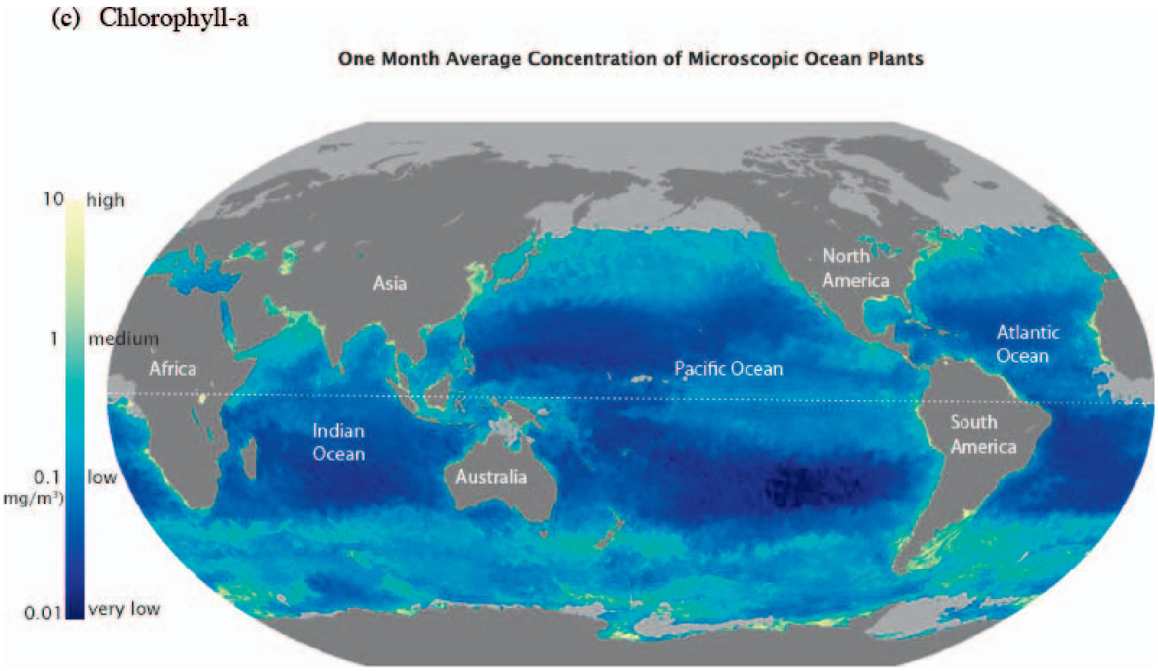

FIGURE 2: continued.

random sample of qualified experts drawn from the university department Web page. Novice participants were recruited using flyers posted on campus and in the surrounding community. After recruitment began on a rolling basis, the final participants in each group were selected more purposively to balance gender representation; female experts were deliberately recruited, and male novices were deliberately recruited through email messages to male-oriented student groups.

Participants were 12 experts (75% presenting male by name and appearance) and 17 novices (35% presenting male). These gender representations were typical of the oceanography department and the general student body of the university. Experts ranged in experience from six to 30 years beyond their PhD. One-third reported working with satellite data specifically in their professional work. Novices were primarily first- and second-year undergraduates, with one nonundergraduate participant. None of the novices had taken an oceanography-specific course at either the high school or undergraduate level, with the exception of one participant who took a marine systems class in high school.

Visualizations and Presentation Order

As part of the larger overall study, different levels of internal scaffolding were presented to determine which, if any, were more useful in helping learners make meaning from the data visualization in the absence of external scaffolds. This dictated the randomization of presentation of the stimuli topics and scaffolding levels. Preliminary quantitative analysis of the results of scaffolding effectiveness are discussed in Stofer (2013) and eye-tracking results from the second study are discussed in Stofer and Che (2014). Therefore, the question of which level of scaffolding was most useful will not be discussed further here except as it relates to the methods for presentation of stimuli.

Participants were presented five versions of each of two global satellite data visualizations with the topics Sea Surface Temperature, Sea Surface Temperature Anomaly, or Chlorophyll-a. The five scaffolding versions for each topic included (1) a version with no scaffolding, ostensibly what a scientist would use to communicate with an audience of peers (unscaffolded); (2) a version with geographic labels for six continents and three ocean basins added (geographic scaffolding); (3) a version with measurement units changed to customary U.S. from metric, and titles revised to remove abbreviations and jargon and include "one month average" for the time span (title scaffolding); (4) a version with the rainbow color scheme revised to use a single-hue or dual-hue, divergent (for SST anomaly) color scheme thought to be more culturally relevant and thus, meaningful to a novice (color scaffolding); and (5) a version that included all three types of single-scaffolding, geography, title, and color (full scaffolding).

The particular topics and order of presentation of the topics was randomized by participant. Presentation of visualization topics was counterbalanced for order, so that four experts and six[2] novices total saw the same two topics and 10 stimuli, with half of each participant group seeing one topic of the pair first, and the other half the opposite. Therefore, participants were shown up to 10 total stimuli in the study described here. The third topic for each participant was reserved for use in the subsequent eye-tracking experiments, described in Stofer and Che (2014).

The scaffolding version of the visualization shown was also randomized by presentation in ExperimentCenter™ presentation software (SensoMotoric Instruments, Inc. Boston, MA) see Tables I and II for presentation orders of topics and an example of scaffolding order for one participant).

[2] Only five novices were shown the chlorophyll/SST anomaly pairing.

TABLE I: Presentation order for stimuli.

Presentation Version	Stimulus		Number of Participants	
	First Topic	Second Topic	Expert	Novice
A	Sea Surface Temperature	Sea Surface Temperature Anomaly	2	3
B[a]	Sea Surface Temperature Anomaly	Sea Surface Temperature	2	3
C	Sea Surface Temperature Anomaly	Chlorophyll-a	2	3
D[a]	Chlorophyll-a	Sea Surface Temperature Anomaly	2	3
E	Sea Surface Temperature	Chlorophyll-a	2	3
F[a]	Chlorophyll-a	Sea Surface Temperature	2	2

[a]Conditions B, D, and F were the same topics as A, C, and E, respectively, with the topic order reversed for presentation.

Interviews

Using a semistructured interview format, participants were asked to examine each visualization version and answer open-ended questions on: (1) the visualization topic, (2) the measurement units used, (3) the meaning of the colors, (4) location of extreme high values, (5) season of the year depicted, (6) time span depicted, and (7) location of the equator. Follow-up probes were used for clarification and to ask participants to describe how they arrived at their answers (see Supplemental Material at http://ufdc.ufl.edu/IR00007549/00001 to be supplied for full interview protocol). Interviews took no more than one hour; some expert interviews were significantly shorter if the participant recognized the data from one stimulus to the next of that same topic. Interviews were conducted by the author.

Analysis

Video recordings of the interviews were transcribed by an external service. The author then reviewed all transcripts for accuracy and made initial notes on codes, including beginning to recognize what would become emergent codes. Next, the author compared participant answers against a rubric developed by the author and reviewed by a colleague who has previously published in the area of meaning-making from ocean data visualizations (Phipps and Rowe, 2010; Rowe et al., 2010; see Supplemental Material at http://ufdc.ufl.edu/IR00007548/00001 to be supplied for full rubrics). Preliminary quantitative assessment against the rubrics is discussed in Stofer (2013) and will not be discussed further here.

An initial codebook was developed with the same colleague with experience in meaning-making from these visualizations. Initial codes were developed to reflect themes expected to appear in participant answers (Miles and Huberman, 1994) based on the conceptual framework, previous work, particular scaffolds chosen for the visualizations, and the interview questions themselves. Expected codes from the constructivist tradition used in the conceptual framework were the central tenets of constructivism, namely, prior knowledge and prior experience. Expected codes from the sociocultural aspects of the conceptual framework included evidence of understanding cultural conventions, particularly related to understanding jargon, scientific units, color schemes, and abbreviations, which were scaffolded in some versions of the stimuli. Expected codes from perception were use on the legend, based on placement on the stimulus and understanding of the color schemes, both of which were tested using scaffolding. Expected codes based on prior work and incorporated in the interview protocol were understanding geography, using patterns in the data for meaning-making, understanding time span, and understanding season. Of these, only geography was scaffolded through labels of continents and the equator in some versions of the stimuli (see Table III for these initial codes).

After the initial codebook was developed and while interviews were continuing, the author began to analyze the data using the initial steps of constant comparative coding (Glaser, 1965). That is, codes were applied to the transcripts line by line with the assistance of NVivo software (QSR

TABLE II: Example stimulus order for participant in presentation version C.

Topic	Scaffolding Level[a]	Presentation Order
Sea Surface Temperature Anomaly	Full scaffolding	1
Chlorophyll-a	Color scaffolding	2
Sea Surface Temperature Anomaly	Unscaffolded	3
Chlorophyll-a	Geographic scaffolding	4
Sea Surface Temperature Anomaly	Color scaffolding	5
Chlorophyll-a	Unscaffolded	6
Sea Surface Temperature Anomaly	Geographic scaffolding	7
Chlorophyll-a	Full scaffolding	8
Sea Surface Temperature Anomaly	Title scaffolding	9
Chlorophyll-a	Title scaffolding	10

[a]Scaffolding level was randomized by Experiment Center software.

TABLE III: Initial codes for data analysis and their sources.

Code	Source for Expected Codes
Prior knowledge	Constructivism portion of framework
Prior experience	Constructivism portion of framework
Understanding of main idea	Sociocultural portion of framework
Use of measurement unit	Sociocultural portion of framework
Use of color legend	Sociocultural and perception portions of framework
Understanding of color meaning	Sociocultural and perception portions of framework
Geography	Interview protocol based on prior work and scaffolding
Data pattern use	Interview protocol based on prior work
Time span of visualization presented	Interview protocol based on prior work
Season of visualization presented	Interview protocol based on prior work

International, Burlington, MA). Each transcript's coding was compared to the other transcripts of participants in that expertise group, and coding was compared across expert and novice groups as well. In the course of constant comparison, additional emergent (Bogdan and Biklen, 2007) codes were revealed by participant responses, including comparison of visualizations within the experiment and confusion about the terms *average* and *normal*. As new codes emerged, the author re-reviewed previously coded transcripts for the additional codes, and compared coding again within and across expertise groups. This iterative process continued until all interviews were complete and the author had applied all expected and emergent codes to all transcripts and no further codes emerged. Once coding was completed, the author grouped codes into categories and themes. The final codebook, including categories and themes, was established in consultation with the same colleague who reviewed the rubric and consulted on the initial codebook; the author completed the full coding of transcripts. Finally, the author determined frequencies of codes for participants overall and by expert and novice population. Frequencies were also computed by group for individual topics, as applicable, and in particular, for main idea, time span, and season, for which participant answers were likely to vary due to the difference in data presented.

RESULTS

Coding

The following qualitative themes were identified in the data. First discussed are the expected themes, created prior to data collection, based on the conceptual framework and interview protocol. Emergent codes that were identified during analysis of participant responses are discussed next.

Expected Themes

Prior Knowledge

All 12 experts reported that their knowledge of how to make meaning of the visualizations came from their graduate studies or professional expertise, though it built on fundamental science studies. As Lindsey[3] said, "Thinking of [the visualizations] in an oceanography context . . . [came] from grad school, but it's all making sense from basic information on seasons learned from a very early age." All novices did report learning about units of measure, seasons, or temperature previously; those that could remember a specific time period (82%) reported learning Fahrenheit and Celsius in elementary (four participants), middle school (six participants), or high school (four participants). By design, however, novice participants lacked the graduate-level study of oceanography of the experts.

Prior Experience

One novice (6%) had seen visualizations such as those shown here in El Niño and hurricane reports in science classes, on TV news, and on the Internet. Overall, novices mentioned having seen visualizations similar to the stimuli here in the news, in a certain science or geography class, with a parent, on weather forecasts, or from a particular teacher. None of them, however, reported extensive experience with these, and five (29%) said they had never seen a visualization like this (see Table IV).

All experts had seen visualizations like these. They said, "Do you know how many times I've seen this image?" (Janet) or "part of [my job] is to develop algorithms that will help us to get better estimates of [this data] from satellites," (Ray) or "I've made those measurements ... I've been working with this data for 30 years" (Charlie). Over half (58%) also produce imagery to visualize their data and are experienced in making choices for representation. Seventy-five percent of the experts teach or taught about visualizations.

Geography

Both groups were able to accurately point out the equator when asked to do so, and all participants reported knowing the equator was in the "middle" or "center" of Earth from a very young age. Experts named many specific ocean features, including Baffin Bay, Labrador Sea, the Gulf Stream, South Pacific gyre, Pacific warm pool, the Kuroshio and Humboldt currents, while novices generally named land features. They mainly named continents, a few countries, and some U.S. state names. One novice (6%) admitted making use of the geographic labels when present to name the specific ocean basins: "Notice how I am identifying them now by their names, because I didn't know that this was the Indian Ocean and Atlantic Ocean before. Because now . . . they're labeled." (Samantha, clinical interview).

[3] All names are pseudonyms.

TABLE IV: Frequencies of codes for all participants and by expertise.

Code	Novices (%) $N = 17$	Experts (%) $N = 12$	Overall (%)
Prior knowledge from graduate school	0 (0)	12 (100)	12 (41)
Prior experience with visualizations	12 (71)	12 (100)	24 (83)
Use of title on first stimulus	6 (35)	9 (75)	15 (52)
Use title to confirm main idea	3 (18)	8 (67)	11 (38)
Use of color legend on first stimulus	14 (82)	11 (92)	25 (86)
Unfamiliar geography	4 (24)	0 (0)	4 (14)
Alternate main idea—Unscaffolded SST[a]	6 (50)	0 (0)	6 (21)
Alternate main idea—Unscaffolded SST[b] Anomaly	4 (33)	0 (0)	4 (14)
Alternate main idea—Unscaffolded Chlorophyll[b]	11 (88)	0 (0)	11 (38)
Used data pattern to understand main idea	9 (53)	10 (83)	19 (66)
Satellite source of data	9 (53)	12 (100)	21 (72)
Satellite cause of missing data	0 (0)	8 (67)	8 (28)
Correct time span of unscaffolded stimulus—SST[c]	0 (0)	0 (0)	0 (0)
Correct time span of unscaffolded stimulus—SST Anomaly[cd]	0 (0)	2 (50)	2 (15)
Correct time span of unscaffolded stimulus—Chlorophyll[ce]	0 (0)	1 (20)	1 (8)
Correct season of stimulus—SST[a]	2 (17)	3 (38)	5 (25)
Comparison of stimuli within experiment	17 (100)	11 (92)	28 (97)
Explicit reference to same data shown previously within experiment	11 (65)	11 (92)	22 (76)
Confusion about average vs. normal	6 (35)	4 (33)	10 (34)

[a]For novices, $n = 12$; for experts, $n = 8$.
[b]For novices, $n = 11$; for experts, $n = 8$.
[c]For novices and experts for this question, n varies due to randomization of stimulus presentation. Some participants saw scaffolded versions of the stimulus title before they saw unscaffolded versions and thus, were not judged on accuracy in that case.
[d]For novices, $n = 9$ (of 11) saw an unscaffolded version first; for experts, $n = 4$ (of 8) saw unscaffolded first.
[e]For novices, $n = 6$ (of 11) saw an unscaffolded version first; for experts, $n = 5$ (of 8) saw unscaffolded first.

While no experts said the geographic focus of the visualization was strange, 24% of novices reported this was an unfamiliar depiction of Earth, centered on the Pacific Ocean basin. While all eventually oriented to the globe, they did not do so immediately. As they caught on, their answers changed: "how much heat is given out in different areas of the world. *Oh, those are the oceans*, like how the temperature of *the water* is in different areas" (Emma). The novice, Brad, who participated in his state geography bee and indicated the most extensive geography experience, thought that the land was the ocean and vice versa for the first fifteen minutes of his interview.

Title, Color Legend, and Measurement Unit Use

Only 35% of novices mentioned they used the title of the visualization when looking to answer the question of main idea on the first of the ten stimuli, while 75% of experts did. Novice Brad even asked, "[do] the words above have any relation to the image?" All participants used the title for evidence by the end of the interview, though they were never prompted to do so by the interview questions.

In addition, 17% of the novices reported failing even to notice the title or legend upon first viewing. When one novice, Veronica, was asked to report the measurement unit for the SST visualization, following questions on both the main idea and the color, she replied "K" (Kelvin). However, she later noticed, "Oh, this one is C. Oh, it shows the unit here! I found it. It should be this," laughed, and pointed to the key, with the units of C marked. Gina remarked, "I didn't even notice the [title] because I've been concentrating so much on pictures." Ivan named the topic of the first stimulus shown to him as simply "water . . . [because] I'm just seeing where the continents are," going on to say the continents were shaded gray while the rest of the visualization was shaded blue, green, and red. Over the course of all stimuli presented, 66% of the experts also reported using the title to confirm their impressions of the main idea based on the data patterns, while only 18% of novices did. Of those, one simply reported that the colors were in the ocean as her comparison with the title.

More of each group of participants (novices 82%, experts 92%) used the color legend than the title to make meaning for the first stimulus. However, in both groups, not all participants explicitly reported of use of the color legend on the first stimulus (82% of novices and 92% of experts), even though all were asked specifically what the colors meant.

Main Idea

For the sea surface temperature visualizations, six of the 12 novices that were shown this topic offered 10 total alternatives to the true main idea of the visualizations. Three references (30%) confused heat and temperature, one (10%) mentioned simply geography, 20% were climate, and 40% of answers mentioned temperatures without referencing the ocean, including one that mentioned the temperature inside the Earth. For the SST anomaly stimuli, even when they noticed the title, the concept of anomaly was unfamiliar, especially since previous stimuli may have shown absolute temperature using the same colors but in different patterns. Participants offered ideas such as depth, currents, and fishing in addition to incomplete ideas about anomaly or confusing heat and temperature when the scaffolded title was not present, the title was not fully understood, or the title was not noticed. Chlorophyll presented the most confusion, reflected in the number of participants offering alternative ideas (88%). *Depth* was mentioned three times; "measuring a substance in the water," *dissolved oxygen/oxygen production*, and *mercury* were each mentioned twice. Three references were made to life in the ocean, one instance each of *harvest*, *fishing*, and *life*. Twice mentioned each were *Climate, rain*, "currents or wind direction for water," *water*, "something to do with where sunlight hits," and "salt or minerals" were each mentioned once.

Data Pattern Use

Novices were less likely than experts to use the patterns of data presented as support for their answers to the question of main idea. About half of the novices (53%) reported only that the colors were in the ocean rather than on land as their only evidence of using the patterns when they were prompted to justify answers based on the data instead of the title or legend. Novice Mikayla said, "I guess that it's the oceans because that's what it's colored in, but I wouldn't know it's chlorophyll." On the other hand, only 17% of experts used only the location of the data in justifying their answers; the other 83% used the distribution of the colors as evidence. As expert Mark explained, "From an oceanographic point of view it makes sense. . . . It shows that the plant life is not distributed evenly through the ocean; there's higher concentrations in the margins and some of the upwelling areas." Nine experts (75%) spontaneously mentioned the continuity of the data was only possible through the use of satellite data; the other three confirmed it was satellite data when asked the source. Only 53% of novices reported that the data was from satellites, only when asked, and often tentatively; one first mentioned that he thought machines inside the Earth measured the temperatures. Allison said, "I don't know how they would do that unless from space they took a picture . . . . I feel we're probably technologically advanced enough to do that, but I don't know."

However, neither group consistently noted a lack of data at the poles due to satellite reliance on visible light. Of the novices, only 29% recognized the poles as areas where no measurements were conducted, though they could not say why that was the case. As Ivan said, it's as if "there was a strict line where they stopped measuring the water." An additional 47% thought the light gray coloring in the stimuli were areas of ice coverage, failing to notice that there were light gray areas near the equator in the ocean that could not be ice. While half the experts noted explicitly there was a lack of satellite coverage depicted in the light gray, and another 25% explained these areas as simply "no data," only two (17%) mentioned the seasonality and lack of light as the reason for the lack of satellite coverage.

Time Span and Season

Novices were overall not familiar with what the concept of a time span depicted by a visualization of this type could mean; five (29%) voiced clarifying questions to that effect. Of those five, two (12%) asked if the interviewer meant time of day or time of year, and three (18%) thought the question was asking about a geological time or era. Three other novices (18%) thought the unscaffolded visualizations depicted instances in time, including one who reasoned that it was the case because the visualization was not itself animated. Three (18%) answered simply that they did not know.

No expert voiced trouble with understanding the question. However, neither group of participants was able to judge the time period very accurately based only on the data presented. If the title did not specify "one month average," no participants answered this accurately on the SST visualizations. Similarly, for SST anomaly and chlorophyll visualizations, no novices answered one month for unscaffolded stimuli. Only two of four experts (50%) correctly identified one month in unscaffolded SST anomaly stimuli, and one of five experts (20%) correctly identified the time span in the unscaffolded chlorophyll visualizations.

Similarly, for both groups, season was difficult to judge from data alone. While both groups recognized the influence of the sun on the seasons, even several of the oceanographers failed to recognize the influence of the specific heat of the ocean, which delays warming and cooling relative to warming and cooling of the air. Only two of the 12 novices (17%) and three of eight experts (38%) that viewed the SST visualizations were able to correctly identify the depicted season.

Emergent Codes

In addition to examining participant meaning-making through the use of codes aligned with the research goals and the interview questions, two additional themes emerged from the participant responses.

Comparison

Although the stimuli were presented sequentially and the participants were not prompted whether or not to consider stimuli shown previously, many spontaneously commented about the similarity or difference of a particular stimuli to one they previously saw in the interview. All but one of the experts (92%) and all novices compared later visualizations with those they were asked about earlier, either noting similar or different colors or patterns. Almost all (92%) of the experts and more than half (65%) of novices explicitly mentioned that they recognized one or more as the same data as previous stimuli.

Average vs. Normal

While some jargon was specifically removed (anomaly, SST, Cholorophyll-a) from the titles of the visualizations, in depicting the time represented in the visualization, confusion around the words *average* (word used in the scaffolded title), *typical, normal,* and *(un)usual* emerged as the interviewer tried to probe participants' meaning-making. Novice Linda stated, "I think of 'difference from average' as normal since temperatures change," reflecting her relative lack of academic experience with the concept. Four novices (24%) explicitly stated similar confusion about what average and normal meant, and two others (12%) expressed that they couldn't answer because they didn't know what average or normal looked like for the stimuli. On the other hand, four experts (33%) talked about disciplinary confusion over the ambiguity of the word *average*: "we always argue about what (pauses) it means to be a one month average. Whether it's the average of every data point in that specific month or if it's the average of all months, (pauses) and all instances of that month, over the lifetime of the satellite," said Brent.

CONCLUSIONS AND IMPLICATIONS

In our experiments, the scaffolded versions—with added geographic labels, more familiar titles, measurement units, and colors—were overall more meaningful to novices as evidenced by their improved meaning-making on these versions. However, novices also improved their meaning-making over the course of the interview through repeated questioning that may have drawn their attention to particular details and asked them to think about the visualizations in ways they had not before. While the interventions did not help novices to completely match the scientific understanding compared to expert scientists based on the meaning-making and confusion codes, the scaffolded versions were more understandable to novices than the unaltered academic versions, consistent with previous work (Phipps and Rowe, 2010; Stofer, 2013; Stofer and Che, 2014).

By its nature, the study had a relatively small number of participants. The participants also demonstrated little demographic diversity based on presentation, reflective of the region of the country from which they were drawn. However, given the entrance requirements to the university, namely, four years of high school science, and the general prevalence of these types of visualizations during the lifetimes of the majority of the novice participants, it seems likely that a great number of members of the public would face similar struggles when trying to interpret unscaffolded or similarly scaffolded visualizations of this type. This is also consistent with previous findings (Keehner et al., 2008; Hegarty et al., 2009; Canham and Hegarty, 2010; Fabrikant et al., 2010; Phipps and Rowe, 2010).

Participants explicitly used prior knowledge and experience to make sense of the visualizations, an important part of learning and meaning-making from the constructivist's perspective (Roschelle, 1995). On the other hand, use of specific elements of visualizations is an example of an enculturated skill, generally learned either through schooling or professional work, demonstrated ably here by the expert participants. Many of the novices did not spontaneously look for either a title or legend, consistent with this view and with previous research with data visualizations (Phipps and Rowe, 2010). Further, even when they reported making use of the elements, the novices did not always arrive at correct answers. This suggests that internal scaffolding of the visualization elements may not be sufficient if the elements are not made salient to the learner. Further, scaffolding on how to obtain information from the elements to make judgments about the data may also be necessary beyond simply presenting clear and obvious supporting information. Comparison by presenting side-by-side visualizations may scaffold this meaning-making, but care must be taken to ensure comparisons do not conflate elements such as the color scale. Evidence here and elsewhere (Rowe et al., 2011) indicates that temperature is often the assumed meaning of the rainbow color scale, further supporting the idea that it is not an ideal representation for a great deal of data visualizations.

Both novices and experts struggled with the depiction of time represented when it was not explicitly spelled out in the supporting information, though experts had some conceptions of the possibilities that could be depicted based on the titles. Also, season confounded a great number of participants in both groups. Novices also generally could not identify the source of the data as from satellites or even from a similar type of spatially and temporally continuous data collection. This suggests that scaffolding can help viewers of all backgrounds absorb given information more readily so that they may focus on making meaning from the data represented.

A recent report on the future of jobs in the United States (Carnevale et al., 2011) finds that science skills, rather than content knowledge, will be most widely applicable to a wide range of both Science, Technology, Engineering, and Math (STEM) and non-STEM careers and thus the broadest swath of learners. While examples in this study center on the ocean visualizations used in this research, the principles for intervention suggested here are likely to help learners of other topics as well, if applied to visualizations of data in those subjects.

Working with data in graphic form, whether from observations, models, or imagination, can draw on a powerful human sense for recognizing patterns and support learners in mastering concepts in national science standards and spatial visualization skills considered essential for geoscientists. However, as the production of these visualizations expands in the professional science realm and they become easily available to the public, facilitators must be sure they present learners with appropriate-level material, either themselves or working in conjunction with visualizers to produce novice and intermediate versions of these representations. Combining appropriate tools with instruction aimed at cognitive style (Kastens et al., 2009; Kastens, 2010) and aimed at different spatial visualization ability (e.g., Titus and Horsman, 2009) can develop visualization interpretation skills and truly harness the potential of these communication tools.

Acknowledgments

The author thanks the Curtis and Isabella Holt Marine Education Award at Oregon State University for a grant to support this work. The author thanks Shawn Rowe for assistance with coding development and assessment of the coding scheme, and Rob Simmon and Jesse Allen of NASA Goddard Space Flight Center for visualization development. The author also thanks Laura Warner, Grady Roberts, and several JGE reviewers and editors for suggestions to improve the manuscript. The SST, climatology, and chlorophyll data for the visualizations were obtained from the Physical Oceanography Distributed Active Archive Center (PO.-DAAC) at the NASA Jet Propulsion Laboratory, Pasadena, CA. http://podaac.jpl.nasa.gov.

References

Abelson, B. 2013. Creating a metric for news apps. Available at https://github.com/abelsonlive/_site_/blob/master/_posts/2013-03-18-A-Metric-For-News-Apps.md (accessed 10 June 2016).

Barthel, C. 2010. Understanding visitor interactions with complex visualizations on spherical display systems in informal learning environments [M.S. thesis]. Corvallis: Oregon State University.

Bogdan, R., & Biklen, S.K. 2007. Qualitative research for education: An introduction to theories and methods, 5th ed. Boston, MA: Pearson A and B.

Breslow, L.A., Ratwani, R.M., and Trafton, J.G. 2009. Cognitive models of the influence of color scale on data visualization tasks. *Human Factors*, *51*(3):321–338.

Canham, M., and Hegarty, M. 2010. Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, *20*(2):155–166. Available at http://doi.org/10.1016/j.learninstruc.2009.02.014 (accessed 3 June 2016).

Carnevale, A.P., Smith, N., and Melton, M. 2011. STEM: Science, technology, engineering, mathematics. Washington, DC: Georgetown University Center on Education and the Workforce.

Chang, H.-Y., and Linn, M.C. 2013. Scaffolding learning from molecular visualizations. *Journal of Research in Science Teaching*, *50*(7):858–886. Available at http://doi.org/10.1002/tea.21089 (accessed 3 June 2016).

Cid, X.C., Lopez, R.E., and Lazarus, S.M. 2009. Issues regarding student interpretation of color as a third dimension on graphical representations. *Journal of Geoscience Education*, *57*(5):372–378. Available at http://doi.org/10.5408/1.3559675 (accessed 3 June 2016).

Committee on the Support for Thinking Spatially. 2006. Learning to think spatially. Washington, DC: National Academies Press.

Conroy, E. 1998. The symbolism of color (originally published 1920). Kila, MT: Kessinger.

Cook, M.P., Wiebe, E.N., and Carter, G. 2008. The interpretation of cellular transport graphics by students with low and high prior knowledge. *International Journal of Science Education*, *30*(2):239–261.

Driver, R. 1995. Constructivist approaches to science teaching. *In* Steffe, L.P. and Gale, J.E., eds., Constructivism in education. Hillsdale, NJ: Erlbaum, p. 385–400.

Edelson, D.C., and Gordin, D. 1997. Creating science learning tools from experts' investigation tools: A design framework. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, Oakbrook, IL, March 21-24, 1997.

Fabrikant, S.I., Hespanha, S.R., and 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. Available at http://doi.org/10.1080/00045600903362378 (accessed 3 June 2016).

Fabrikant, S. I., and Lobben, A. 2009. Introduction: Cognitive issues in geographic information bisualization. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *44*(3):139–143. Available at http://doi.org/10.3138/carto.44.3.139 (accessed 3 June 2016).

Glaser, B.G. 1965. The constant comparative method of qualitative analysis. *Social Problems*, *12*(4):436–445. Available at http://doi.org/10.2307/798843 (accessed 3 June 2016).

Goodchild, M.F., and Janelle, D.G. 2010. Toward critical spatial thinking in the social sciences and humanities. *GeoJournal*, *75*(1):3–13. Available at http://doi.org/10.1007/s10708-010-9340-3 (accessed 3 June 2016).

Haley Goldman, K., Kessler, C., and Danter, E. 2010. Science on a sphere: Cross-site summative evaluation. Edgewater, MD: Institute for Learning Innovation. Available at http://www.oesd.noaa.gov/network/SOS_evals/SOS_Final_Summative_Report.pdf (accessed 3 June 2016).

Hegarty, M. 2011. The cognitive science of visual–spatial displays: Implications for design. *Topics in Cognitive Science*, *3*(3):446–474. Available at http://doi.org/10.1111/j.1756-8765.2011.01150.x (accessed 3 June 2016).

Hegarty, M. 2013. Cognition, metacognition, and the design of maps. *Current Directions in Psychological Science*, *22*(1):3–9. Available at http://doi.org/10.1177/0963721412469395 (accessed 3 June 2016).

Hegarty, M., Smallman, H.S., Stull, A.T., and Canham, M.S. 2009. Naïve cartography: How intuitions about display configuration can hurt performance. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *44*(3):171–186. Available at http://doi.org/10.3138/carto.44.3.171 (accessed 3 June 2016).

John-Steiner, V., and Mahn, H. 1996. Sociocultural approaches to learning and development: A Vygotskian framework. *Educational Psychologist*, *31*(3):191–206.

Kastens, K.A. 2010. Commentary: Object and spatial visualization in geosciences. *Journal of Geoscience Education*, *58*(2):52–57. Available at http://doi.org/10.5408/1.3534847 (accessed 3 June 2016).

Kastens, K.A., Agrawal, S., and 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.

Kastens, K.A., and Ishikawa, T. 2006. Spatial thinking in the geosciences and cognitive sciences: A cross-disciplinary look at the intersection of two fields. In Manduca, C.A., and Mogk, D.W., eds., Earth and mind: How geologists think and learn about the earth. Boudler, CO: Geological Society of America, p. 53–76.

Keehner, M., Hegarty, M., Cohen, C., Khooshabeh, P., and Montello, D.R. 2008. Spatial reasoning with external visualizations: What matters is what you see, not whether you interact. *Cognitive Science*, *32*(7):1099–1132. Available at http://geog.ucsb.edu/∼montello/pubs/med_visualize.pdf (accessed 10 June 2016).

Kozma, R., Chin, E., Russell, J., and Marx, N. 2000. The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning. *The Journal of the Learning Sciences*, *9*(2):105–143.

Light, A., and Bartlein, P.J. 2004. The end of the rainbow? Color schemes for improved data graphics. *EOS Transactions of the American Geophysical Union*, *85*(40):385–391.

Miles, M.B., and Huberman, A.M. 1994. Qualitative data analysis: An expanded sourcebook, 2nd ed. Thousand Oaks, CA: Sage.

Nyhan, B., and Reifler, J. 2015. The roles of information deficits and identity threat in the prevalence of misperceptions. Paper presented at the frank Gathering, Gainesville, FL, February 22-24. Available at http://www.dartmouth.edu/∼nyhan/opening-political-mind.pdf (Accessed June 10, 2016).

Phipps, M., and Rowe, S.M. 2010. Seeing satellite data. *Public Understanding of Science*, *19*(3):311–321. Available at http://doi.org/10.1177/0963662508098684 (accessed 3 June 2016).

Roschelle, J. 1995. Learning in interactive environments: Prior knowledge and new experience. In Falk, J.H., and Dierking, L.D., eds., Public institutions for personal learning: Establishing a research agenda. Washington, DC: American Association of Museums, p. 37–51.

Rowe, S.M., Stofer, K., Barthel, C., and Hunter, N. (2010). Hatfield Marine Science Center Magic Planet installation evaluation findings. Corvallis, OR: Oregon Sea Grant.

Rowe, S.M., Stofer, K., Bullick, S., and O'Brien, S. 2011. GEO SSI project Phase 1 evaluation results. Corvallis, OR: Oregon State University.

Serway, R.A., and Faughn, J.S. 2009. College physics, 8th ed. Belmont, CA: Brooks/Cole.

Steffke, C.L., and Libarkin, J.C. 2012. Guiding symbology and display selection to produce more effective images for conveying information. Poster presented at the Geological Society of America, Charlotte, NC, November 4–7.

Steffke, C.L., and Libarkin, J.C. 2013. 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, October 27–30. Available at https://gsa.confex.com/gsa/2013AM/finalprogram/abstract_232424.htm (accessed 3 June 2016).

Stofer, K.A. 2013. Visualizers, visualizations, and visualizees: Differences in meaning-making by scientific experts and novices from global visualizations of ocean data (doctoral dissertation). Corvallis: Oregon State University.

Stofer, K.A., and Che, X. 2014. Comparing experts and novices on scaffolded data visualizations using eye-tracking. *Journal of Eye Movement Research*, *7*(5):1–15.

Taber, M.R., Ledley, T.S., Lynds, S., Domenico, B., and Dahlman, L. 2012. Geoscience data for educational use: Recommendations from scientific/technical and educational communities. *Journal of Geoscience Education*, *60*(3):249–256. Available at http://doi.org/10.5408/12-297.1 (accessed 3 June 2016).

Titus, S., and Horsman, E. 2009. Characterizing and improving spatial visualization skills. *Journal of Geoscience Education*, *57*(4):242–254. Available at http://doi.org/10.5408/1.3559671 (accessed 3 June 2016).

Tversky, B. 2014. Visualizing thought. *In* Huang, W. ed., Handbook of human centric visualization. New York, NY: Springer, p. 3–40. Available at http://link.springer.com/chapter/10.1007/978-1-4614-7485-2_1 (accessed 3 June 2016).

Vygotsky, L. 1978. Mind in society: The development of higher mental processes. Cambridge, MA: Harvard University Press.

Wood, D., Bruner, J.S., and Ross, G. 1976. The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2):89–100. Available at http://doi.org/10.1111/j.1469-7610.1976.tb00381.x (accessed 3 June 2016).