ABSTRACT

MAUDLIN, LINDSAY CATHERINE. An Evaluation of a Climate Decision Support System: An Eye-Tracking Study. (Under the direction of Dr. Karen S. McNeal and Dr. Walter A. Robinson).

A web-based climate decision support system (DSS) designed for foresters in the Southeast United States was evaluated to measure the overall DSS usability and to determine if user characteristics (e.g., age, gender, education) impact usability. This was accomplished by asking users to navigate through the web-based climate DSS to complete tasks and answer multiple choice questions about the climate information they interact with. They were also asked for basic demographic information and about their experiences using climate information. The evaluation utilized eye tracking, a technology that determines where, when, and for how long a user fixates or focuses attention on a particular element on a computer screen. Eye-tracking data can provide insight into the paths users take to complete the tasks and answer the questions, which website design elements are most salient, and which elements are never viewed. The data were collected in two phases: the first was exploratory and prompted changes to be made to the DSS, and the second aimed to determine how those changes influenced the usability of the DSS. This work explores three main areas: the differences between males and females; the differences between experts and novices; and how the usability of the web-based climate DSS changed between the two data collection phases.

Males and females demonstrated different eye-tracking patterns throughout the study. In general, males tended to focus their attention on the maps and climate data, while females tended to focus more attention on other elements of the DSS such as buttons that could be toggled off and on, color bars, and text boxes. Additionally, when females did look at the maps and climate data, they did not look at the most important aspects of the maps and climate data for answering

the related questions. This then impacted their success in answering the questions correctly. Overall, males answered more questions correctly than did females, likely because males were looking at the relevant data.

In this study, experts were defined as forestry professionals and novices were defined as undergraduate students. The expected expert-novice differences (e.g., experts answered more questions correctly) were convoluted by the influence of age on DSS usability. The expert population had a higher average age than the novice population, and they also performed worse than the novices. This unexpected results is likely due to the impact of age on website usability. In addition, experience using climate data was not an accurate predictor of performance. The lowest performers had experience using climate data, and not all of the highest performers did. While there were some differences in the eye-tracking patterns between experts and novices in this study, likely due to experience and content expertise, age was the most important factor in the eye-tracking data.

The usability changed between the two phases of the study. In general, the efficiency improved after changes were made to the DSS. The overall effectiveness decreased, with an exception where users in the second phase out-performed those in the first phase during one part of the study. This unexpected finding suggested users might have been cognitively overloaded, thereby decreasing their performance.

An Evaluation of a Climate Decision Support System: An Eye-Tracking Study

by
Lindsay Catherine Maudlin

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Marine, Earth, and Atmospheric Sciences

Raleigh, North Carolina

2018

APPROVED BY:

Dr. Karen S. McNeal Committee Co-Chair	Dr. Walter A. Robinson Committee Co-Chair
Dr. Ryan Boyles	Dr. Nicole Lee

DEDICATION

To my grandparents, Earl, Kathy, Pete, and Velma.

Thank you for passing on your grit.

BIOGRAPHY

Lindsay C. Maudlin was born and raised in Hutchinson, Kansas. She loved weather from a young age, but she also loved animals and wanted to be a veterinarian when she grew up. Her dreams changed as she learned more about meteorology and the possible career paths in that field, and by the end of her freshman year of high school, Lindsay had decided to become a meteorologist. After high school, Lindsay went to Hutchinson Community College where she earned an Associate in Arts and then transferred to The University of Oklahoma where she earned her Bachelor of Science in Meteorology. During her undergraduate years, Lindsay studied abroad at Monash University in Australia, where she developed an interest in atmospheric chemistry. This interest led her to The University of Arizona where she earned her Master of Science in Atmospheric Science and completed one year of the PhD program while conducting research on the physicochemical properties of atmospheric aerosols through fieldwork and instrumentation. During the first year of her PhD, Lindsay was also a fellow in the NSF GK-12 program. She was paired with a gifted education teacher and developed and taught various STEM lesson plans for 5th and 6th grade students. This experience was an opportunity for Lindsay to develop her teaching skills, and it also introduced her to education research, ultimately leading her to North Carolina State University to complete her PhD in Atmospheric Science with a research focus on Geoscience Education Research and Geocognition.

In addition to her academic pursuits, Lindsay has continued to pursue ballet and long-distance running. She also enjoys reading, music, traveling, hiking, and being a cat mother to a cheeky Siamese named Canberra.

ACKNOWLEDGMENTS

I am the product of a public school education, kindergarten through PhD, and also the recipient of state and federal funding for my graduate studies. I am a testament to the good that adequate funding for public schools and science can do, and I know that I would not be here if not for such funding. As such, I would like to thank the taxpayers who have made this possible.

Along my public school journey, I had many wonderful teachers and professors who taught, encouraged, and mentored me. A few in particular were especially influential in my journey:

- John Brown, Hutchinson High School
- Rusty Hilst, Hutchinson High School
- Jake Miller, Hutchinson High School
- Lynn Anderson, Hutchinson Community College
- Dr. Mike Biggerstaff, The University of Oklahoma
- Dr. Ave Arellano, The University of Arizona
- Dr. Liz Ritchie, The University of Arizona
- Dr. Randy Richardson, The University of Arizona

During my time at The University of Arizona, I spent a fair amount of time in K-12 classrooms through various science outreach activities, as a GK-12 Fellow, and as an instructor for a weather and climate summer camp. Four teachers generously shared their classrooms and students with me, and I am grateful for the lessons they taught me:

- Michele Thelen
- Beth DeWeerdt
- Lissa Bouwens

- Shari Popen

My former students and my future students have been an endless source of motivation to earn my PhD. Teaching has been the best part of my graduate school experience, and I cannot wait to teach again.

My graduate studies at NCSU have been funded by the Southeast Climate Adaptation Science Center's Global Change Fellowship and the Provost's Fellowship. These fellowships have allowed me to focus on and complete my research on a shorter than usual timeline.

In addition to funding my graduate studies, the Southeast Climate Adaptation Science Center and its staff have provided me with opportunities to meet and get to know graduate students from other departments, attend and organize seminars, and participate in various stakeholder meetings, all of which have grown my professional networks and helped me forge new friendships.

The Marine, Earth, and Atmospheric Sciences staff have saved the day on many occasions, but two women in particular have been instrumental in my success:

- Jackie Sullivan
- Meredith Henry

My graduate advisors and PhD committee members have been influential in my growth as a scientist, and I am grateful for their guidance, encouragement, and wisdom:

- Dr. Armin Sorooshian, The University of Arizona, MS Committee Chair
- Dr. Karen S. McNeal, Auburn University, PhD Committee Co-chair
- Dr. Walter A. Robinson, North Carolina State University, PhD Committee Co-chair
- Dr. Ryan Boyles, Southeast Climate Adaptation Science Center, PhD Committee Member

- Dr. Nicole Lee, North Carolina State University, PhD Committee Member

My peers and academic siblings have commiserated with, taught, and encouraged me throughout my graduate school years. I am grateful for the late night study sessions, coffee runs, distractions, laughs, and sometimes tears shared with them. I've enjoyed the friendships that formed as a result of these experiences, particularly those with Taylor Shingler and Nick Soltis who have been the best academic siblings and friends I could have asked for.

As I walked through the desert that has been my last three years, God has provided for all my needs and parted the Red Seas in my life by placing particular people along my path. As I have personally experienced Ecclesiastes 12:12, which states, "Of making many books there is no end, and much study is a weariness of the flesh," I have been encouraged by many people, but there are a few I'd like to mention in particular:

- My Medora, North Swan, and Providence families
- My run club friends: Amber Rhyne, Allen Boone, and Brandon Andrews
- My clothespins and fellow ALA SGS staff: Rachel Barnes, Jackie Paschang, Heather Williams, Liz Woody, Amy Stotts, and Stacy Edwards
- The Mikes: Mike Madden and Mike Pirhalla
- Fellow OU Alumni: Stefan Rahimi and Stephen Castleberry
- My dearest high school friends: Tara Jones, Cindy Mood, Lauren Edwards, and Bianca James
- My best friends and their families: The Silsbees, Warnkens, and Hemmans
- My family
- My MVPs: Canberra, Gina Long, and Ian Smith

The six most important people responsible for getting me here through their unconditional love, support, sacrifices, collective wisdom, and top-notch genetics are Craig, Linda, Earl, Kathy, Pete, and Velma. I would not be where I am now if not for their positive impacts on my life. This is as much theirs as it is mine.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: Introduction

Loblolly pine trees are a critical component of the forestry industry in the Southeast United States. They comprise nearly 20 million acres of forests in this region and provide economic and ecological services (PINEMAP.org). These trees serve as a major carbon sink in the United States, but they are also sensitive to variable and changing climates. Due to their importance and susceptibility to climate risks, the Pine Integrated Network: Education, Mitigation, and Adaptation Project (PINEMAP) was proposed and funded through the USDA National Institute of Food and Agriculture in 2011. The project aims to combine research, extension, and education to help foresters in this region manage loblolly forests in a way such that they will increase carbon sequestration, resiliency, and sustainability.

PINEMAP ultimately led to the development of a climate decision support system (DSS), the PINEMAP DSS, which provides foresters in the Southeast United States with historical climate data as well as climate projections. The data are organized into different categories (e.g., Environment, Establishment, Management, and Production) and further divided into different tools that focus on one variable. After the PINEMAP DSS was developed, it underwent beta testing. The subsequent evaluation, which uses eye tracking, is presented here.

Literature Review

As the impacts from climate change begin to be observed, the need to disseminate climate information to decision-makers for adaptation purposes grows and, in some cases, is legally required (USDOI, 2009). When climate information is shared, special consideration must be given to the process by which climate information is selected, used, and explained such that the decision-makers fully understand the information they are using to make decisions for ecological purposes (Snover et al., 2013), as well as how it is visualized, designed, and tailored for specific decision-makers and the layperson (Daron et al., 2015; Harold et al., 2016). Despite the body of literature that acknowledges the need for readily understood and accessible climate information for decision-makers, there is limited literature on how individuals and groups interpret the climate information they are presented (McInerny et al., 2014; Daron et al., 2015).

Lessons learned from the field of communication can be applied to the practice of disseminating climate information to decision-makers. Audiences bring many factors (e.g., political affiliation and ideology, education level, experience with major weather events, proximity to climate change victims, etc.) to the table that can hinder their ability to accept the science and take action (Borick and Rabe, 2008; Borick and Rabe, 2010; McCright and Dunlap, 2011; Hart and Nisbet, 2011; Kahan et al., 2012; Pearson and Schuldt, 2015). While addressing all of these factors at once would be impossible, being aware of their impacts on an audience's cognitive, affective, and conative states is an important part of the communication process. Potential ways to increase the likelihood of an audience accepting the science and taking action include: tailoring the message for the intended audience (Severin and Tankard, 2001); allowing for interaction between the information and the audience (Metzger, 2009); and creating space for audience participation and feedback in the communication process (Nisbet and Scheufele, 2009).

Decision support systems (DSSs) are one way that climate information can be tailored to decision-makers in a specific discipline, and there have been requests for such DSSs to be developed (Mahoney et al., 2003). DSSs are, in general, collections of information (i.e., scientific research, analyses of information, and factors important to the decision-making process) located in a common space that decision-makers can access and consult when making a decision. They span across many disciplines and are becoming more common (Pyke et al., 2007). Unfortunately, the efficacy of a DSS is not often studied and there is often a gap between the information provided by a DSS and the needs of the decision makers using it (Pyke et al., 2007). Furthermore, DSSs are often housed on websites, making them interactive (Metzger, 2009; Mysiak et al., 2005) while also introducing the need to study the usability of web-based DSSs.

Website usability has been studied in great detail across many disciplines, and the lessons learned through this body of research can perhaps shed some light on stumbling blocks or usability concerns for DSSs that are often shared through websites. Website usability can be influenced by certain user demographics. For instance, older adults tend to be less accurate and take longer to complete tasks (Romano Bergstrom et al., 2013), and age has been found to impact spatial ability, mental model accuracy, and usability through loss in cognitive skills (Wagner et al., 2014; Grahame et al., 2004). When gender is considered, males and females have different design preferences for websites, and the male dominated information technology industry could potentially be a stumbling block for women using websites designed by males (Moss et al., 2006). Additionally, men and women differ in what they consider to be noticeable and appealing through self-reported surveys, but the coordinating eye-tracking data do not show significant differences in the eye movements of males and females, leading to the need for additional gender studies in website usability (Djamasbi et al., 2007).

Eye tracking is a quantitative research method that determines where, when, and for how long a participant looks at a particular element on a computer screen or otherwise placed within the line of sight. This method has existed for more than 100 years (Jacob and Karn, 2003) and has been used in many different applications (Holmqvist et al., 2011; Drost et al., 2015; Wilson et al., 2016). Eye tracking is often used in website usability studies in order to provide higher resolution data than observable behaviors such as scrolling or mouse clicks (Bojko, 2006; Loyola et al., 2015). The limitations of eye tracking are that it only determines what the user is looking at instead of determining why the user looks at something or whether or not the user likes that feature (Velasquez, 2013), and the analysis can be very time intensive (Jacob and Karn, 2003). Through this technology, researchers have learned that not only do website usability problems correlate well with specific eye-tracking patterns, but also to a series of patterns or behaviors that users exhibit as they struggle to complete a given task (Ehmke and Wilson, 2007). It is important to note that these studies tend to have small sample sizes (<40) (Jacob and Karn, 2003).

Motivation and research questions

There is a need to successfully disseminate accurate climate information to decision-makers as the urgency for mitigation and adaptation planning grows, but presently, there is little information on the efficacy of the existing dissemination methods. Specifically, the gaps identified in the literature are: how individuals and groups of decision-makers interpret the climate information they are provided; the efficacy of web-based DSS tools designed for a specific audience; and how user demographics such as age, education, and gender might influence a user's ability to navigate a DSS. Without evaluation, there is no way to determine if the existing methods of disseminating climate information to decision-makers are successful.

This evaluation aims to address the gaps identified in the literature by evaluating a web-based DSS designed for a specific audience – foresters in the Southeast United States. The study will utilize eye-tracking technology to identify aspects of the DSS that should be altered to improve the overall usability and to quantify the differences between the various demographics examined. Furthermore, this work will address the following broad research questions:

1. Are there gender differences in users' ability to successfully navigate the web-based DSS and interpret the information? In what ways are gender differences manifested?

2. Is experience, defined broadly by education level, age, and experience with climate information, an accurate predictor of performance with a web-based DSS? Are there differences between experts and novices, or are there other factors involved?

3. Which design elements distract or confuse users? Through the feedback process, changes can be made to the overall design of the DSS. Do the changes help users better navigate the tools and answer the questions more quickly and/or accurately? Are there other usability concerns to examine?

By answering these questions, this evaluation will contribute to several fields by filling the gap in the literature and serve as the first eye-tracking-focused evaluation of a web-based climate information DSS. Additionally, it will add to the limited body of literature on gender differences and expert-novice differences in website usability within an even more narrow area of climate information websites. Eye tracking is very quantitative, thus adding a more robust evaluation of website usability, particularly between groups, to the existing body of literature.

Methods

PINEMAP DSS

As part of the Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP), the PINEMAP DSS was created to provide climate and regional productivity model data specifically for foresters in the Southeast United States. The tools that comprise the DSS examine the influences that climate and loblolly pines have on each other. The following three tools were chosen for evaluation purposes: Extreme Minimum Temperature (EMT); Summer Precipitation (SP); and Seedling Markets (SM), later renamed the Cold-Tolerant Markets for Nurseries (CTMN).

Study Designs and Populations

There were two phases in the evaluation of the PINEMAP DSS tools. The first phase (Phase 1) was exploratory and was conducted with experts in the forestry field. Eye tracking was used to help determine which aspects of the website impacted a user's ability to successfully use the website. Usability was measured by asking users to complete one task and answer two multiple choice questions for each tool (three tasks and six questions in total). The tasks were designed to utilize certain features of the PINEMAP DSS, one at a time, and to build on each other from the first to the third task in order to measure the usability of individual features. For example, the first task asked the users to input latitude and longitude coordinates, and the last task asked the users to enter the latitude and longitude coordinates and find a drop-down menu to turn-on a specific map layer. User demographics (age, gender, and education level) were also collected to determine if certain characteristics influenced a user's ability to complete the tasks and answer the related questions correctly. Additionally, participants were asked open-ended questions about their experiences using climate data and climate information websites. Based on the initial findings from the first phase, the PINEMAP DSS was updated. The second phase (Phase 2) aimed to determine if the updates made the PINEMAP DSS easier to use, if the statistically significant differences in user performance based on user characteristics from Phase 1 could be replicated with a new group of participants (novices) in Phase 2, and if differences between the participants in each phase exist based on education and experience levels.

The first participant group included 30 people (18 males and 12 females; 21 to 65 years old), and their highest degrees, either completed or in progress, ranged from undergraduate to doctoral degrees (6 undergraduate degrees, 17 master's degrees, and 6 doctoral degrees). These participants were forestry students and professionals, were recruited at a forestry conference in

Durham, North Carolina, and were compensated with their choice of a pre-packaged candy bar or a rain gauge. All participants: had five minutes to freely explore the PINEMAP DSS; had an unlimited amount of time to complete the tasks and answer the related questions for the EMT, SP, and SM tools; and completed a questionnaire (demographics and experience with climate data).

The second participant group included 12 undergraduate students (4 males and 8 females; 18 to 21 years old), and their grade levels ranged from freshmen to juniors (2 freshmen, 8 sophomores, and 2 juniors). These participants were recruited from introductory physical science classes at a large research institution in the southeastern US and compensated with a pre-packaged candy bar. All participants: had five minutes to freely explore the PINEMAP DSS; had an unlimited amount of time to complete the tasks and answer the related questions for the EMT, SP, and CTMN tools; completed a questionnaire (demographics, experience with climate data, and experience with activities linked to developing spatial ability such as scouts, video games, navigation, etc.); and completed a block slicing test (Titus and Horsman, 2009; Ormand et al., 2014) to measure one aspect of spatial ability.

Both phases of this study were carried out under approval through the Institutional Review Board for the Protection of Human Subjects.

Eye-tracking Specifications and Analyses

The eye-tracking hardware used in the evaluations of the PINEMAP DSS consisted of two Tobii X2-60 (60 gaze points per second) eye trackers during Phase 1 and one Tobii TX-300 (300 gaze points per second) eye tracker during Phase 2. The Tobii X2-60 set-ups (portable laptops) were used to collect data remotely (at a forestry conference in Phase 1), and the TX-300 set-up (stationary desktop) was used to collect data in a lab setting (on campus in Phase 2).

Before eye movements were recorded, each participant underwent a calibration process and was seated such that the eyes were ~65 cm from the eye tracker, all to ensure accurate and consistent eye-tracking data. After the eye-tracking data has been collected, the Tobii software provided both qualitative and quantitative outputs for analyses.

The qualitative outputs consist of heat maps, which indicate where and how many fixations occur, and gaze plots, which indicate when and where users look. These products are suitable for visual analysis and general descriptions of user eye movements and behaviors but not for statistical analysis purposes.

The quantitative outputs are calculated based on areas of interest (AOIs), regions within the PINEMAP DSS that are defined by the researcher. For example, one AOI is the region in each of the tools where the user would enter in the latitude and longitude coordinates for a specific location of interest. The AOIs were consistent as possible between Phase 1 and Phase 2 of the study, noting any changes between the two versions of the PINEMAP DSS tools when necessary.

Once the AOIs are defined, three quantitative metrics are then calculated specific to each AOI: the fixation count (FC), which is the total number of times that the user fixates anywhere within the AOI; the time to first fixation (TFF), which is the total number of seconds between the time at which the webpage loads and the time at which the user fixates for the first time within the AOI; and the total visit duration (TVD), which is the total number of seconds over all visits to the AOI that the user fixates within the AOI. These metrics are then exported from Tobii for statistical analyses with the other datasets (questionnaire, tasks and multiple choice questions, and object slicing test for the participants in Phase 2). For quality control purposes, the eye-tracking data (i.e., FC, TFF, and TVD) from participants with low calibration rates (<60%) will not be included in the statistical analyses.

Questionnaire, Tasks/Questions, and Object Slicing Test Analyses

The questionnaire responses will be compiled into a single document. Age, gender, and education level responses will be used for statistical analyses, and the open-ended responses will be used to determine previous experience with climate data and spatial skill-building activities (Phase 2 users only). Users' responses to the multiple-choice questions asked during the task portion of both phases of the study will be graded and given points based on correctness (1 point for a correct response, and 0 points for an incorrect response). Individual scores will be compared to the AOI metrics from the corresponding tools to look for eye movements or behaviors that correspond to answering the question correctly. For users in Phase 2, the object slicing test will be graded (one point for each correct answer, zero points for questions left blank, and a quarter of a point taken off for each incorrect answer). Overall scores for the tasks and multiple-choice questions (the total number of points out of 6) and for the object slicing test (possible scores between -3.75 and 15) will be compared between the genders, age groups, and education levels of the participants to determine if certain user characteristics influence the users' ability to correctly answer the questions.

Statistical Analyses

Simple means, medians, and standard deviations will be calculated for each of the eye-tracking metrics for all participants, for males and females, for each level of education (undergraduate, master's, doctoral) or undergraduate grade level, and for the various age groups. Statistical tests will be conducted to determine if the differences in means between these groups are statistically significant and to determine if any factors can successfully predict a performance outcome. Additionally, tests will be performed to determine if the differences in means between the various groups (age, gender, education, experience, etc.) are significant. These analyses will be completed using a combination of Excel and SPSS:

- Means (Student's t-test) – All metrics
- Medians (Independent-Samples, k-samples) – All metrics
- Correlations (Spearman 2-tailed test) – All metrics
- Distributions (Mann-Whitney U-test) – All metrics
- Factor Analysis – Fixation Count; with Levene's Test (variances)
- Effect Size – All metrics

Limitations

Due to the inherent differences between the forestry students and professionals from Phase 1 and the undergraduate students in introductory physical science courses from Phase 2, the populations are not identical. Additionally, changes were made to the PINEMAP DSS website between Phase 1 and Phase 2 (based on the preliminary findings from Phase 1). These could potentially influence the results that compare metrics between participants in Phase 1 and those in Phase 2.

Spatial ability is a broad skill-set that is best measured through multiple instruments (Titus and Horsman, 2009; Ormand et al., 2014). Due to limited time with each participant, only one instrument was used for the participants in Phase 2, thus limiting the measurement of spatial ability to one aspect of the broader skill-set.

Validity

The tasks and questions specific to the PINEMAP DSS website were written with help from the website developer and reviewed by three climatologists and a geoscience education researcher. Once the tasks and questions were added to the eye-tracking study design, the entire study was reviewed with the help of four graduate students in the geosciences for content validity.

Chapter descriptions

A brief summary of the questions to be answered by each chapter is provided. The population to be studied, the eye-tracking metrics, any additional data, and the analyses to be used are also designated (Table 1).

Gender differences

Statistically significant differences were found between male and female participants from Phase 1 for the SM tool. Do similar differences exist in Phase 2 for the CTMN tool? Can these differences be explained by spatial abilities (measured only in Phase 2 with an object slicing test)? Are there similarities between the males and females in both populations?

Expert-Novice differences

Phase 1 consisted of students and professionals in the forestry business (Experts), and Phase 2 consisted of students in introductory physical science classes (Novices). Controlling for modifications made to the PINEMAP DSS between the two phases, are there any statistically significant differences between Experts and Novices in how they engaged with the DSS PINEMAP website?

Usability

What are the major design obstacles for users? Which features are distracting or confusing? Once changes have been made, are users better able to navigate and use the PINEMAP DSS?

References

Bojko, A., 2006: Using Eye Tracking to Compare Web Page Designs: A Case Study.

Borick, C.P., Rabe, B.G., 2008: A reason to believe: Examining the factors that determine Americans' views on global warming. Issues in Governance Studies, 18, 1-14.

Borick, C.P., Rabe, B.G., 2010: A reason to believe: Examining the factors that determine individual views on global warming. Social Science Quarterly, 91, 777–800.

Daron, J.D., Lorenz, S. Wolski, P., Blamey, R.C., Jack, C., 2015: Interpreting climate data visualizations to inform adaptation decisions. Climate Risk Management, 10, 17-26.

Djamasbi, S., Tullis, T., Hsu, J., Mazuera, E., Osberg, K., 2007: Gender Preferences in Web Design: Usability Testing through Eye Tracking. Americas Conference on Information Systems, AMCIS 2007 Proceedings.

Drost, R., Trobec, J., Steffke, C., Libarkin, J., 2015: Eye tracking: Evaluating the impact of gesturing during televised weather forecasts. Bull. Amer. Meteor. Soc., 96, 387–392,

Ehmke, C., Wilson, S., 2007: Identifying Web Usability Problems from Eye-Tracking Data. British Computer Society, Proceedings of HCI 2007.

Grahame, M., Laberge, J., Scialfa, C.T., 2004: Age Differences in Search of Web Pages: The Effects of Link Size, Link Number, and Clutter. Human Factors, 46, 385-398.

Harold, J., Lorenzoni, I., Shipley, T.F., Coventry, K.R., 2016: Cognitive and psychological science insights to improve climate change data visualization. Nature Climate Change, 6, 1080-1089.

Hart, P.S., Nisbet, E.C., 2011: Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies. Communication Research, 39, 701-723.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., van de Weijer, J., 2011: Eye Tracking: A comprehensive guide to methods and measures. Oxford, 560 pp.

Jacob, R.J.K., Karn, K.S., 2003: Eye Tracking in Human–Computer Interaction and Usability Research: Ready to Deliver the Promises. The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research. Hyona, Radach & Deubel (eds.) Oxford, England.

Kahan, D.M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L.L., Braman, D. & Mandel, G., 2012: The polarizing impact of science literacy and numeracy on perceived climate change risks. Nature Climate Change, 2, 732-735.

Loyola, P., Martinez, G., Munoz, K., Velasquez, J.D., Maldonado, J., Couve, A., 2015: Combining eye tracking and pupillary dilation analysis to identify Website Key Objects. Neurocomputing, 168, 179-189.

Mahoney, J.R., Asrar, G., Leinen, M.S., Andrews, J., Glackin, M., Groat, C., Hobenstein, W., Lawson, L., Moore, M., Neale, P., Patrinos, A., Schafer, J., Slimak, M., Watson, H., 2003: Strategic Plan for the U.S. Climate Change Science Program, Ch. 11 Decision Support Resources Development.

McCright, A.M., Dunlap, R.E., 2011: The politicization of climate change and polarization in the American public's views of global warming, 2001-2010. The Sociological Quarterly, 52, 155-194.

McInerny, G.J., Chen, M., Freeman, R., Gavaghan, D., Meyer, M., Rowland, F., Spiegelhalter, D.J., Stefaner, M., Tessarolo, G., Hortal, J., 2014: Information visualisation for science and policy: engaging users and avoiding bias. Trends in Ecology & Evolution, 29, 148-157.

Metzger, M.J., 2009: The study of media effects in the era of Internet communication. In R. Nabi & M. B. Oliver (Eds.), Handbook of media effects, 561–576.

Moss, G.A., Gunn, R., Heller, J., 2006: Some Men Like It Black, Some Women Like It Pink: Consumer Implications of Differences in Male and Female Website Design. Journal of Consumer Behaviour, 5, 328-341.

Mysiak, J., Giupponi, C., Rosata, P, 2005: Towards the development of a decision support system for water resource management. Environmental Modelling & Software, 20, 203-214.

Nisbet, M.C., Scheufele, D.A., 2009: What's next for science communication? Promising directions and lingering distractions. American Journal of Botany, 96, 1767–1778.

Ormand, C.J., Manduca, C., Shipley, T.F., Tikoff, B., Harwood, C.L., Atit, K., Boone, A.P., 2014: Evaluating Geoscience Students' Spatial Thinking Skills in a Multi-Institutional Classroom Study. Journal of Geoscience Education, 62, 146-154.

Pearson, A.R., Schuldt, J.P., 2015: Bridging climate communication divides: Beyond the partisan gap. Science Communication, 37, 805-812.

PINEMAP (http://climate.ncsu.edu/pinemap_dss/index.php?tool=)

Pyke, C.R., Bierwagen, B.G., Furlow, J., Gamble, J., Johnson, T., Julius, S., West, J., 2007: A decision inventory approach for improving decision support for climate change impact assessment and adaptation. Environmental Science & Policy, 10, 610-621.

Romano Bergstrom, J.C., Olmsted-Hawala, E.L., Jans, M.E., 2013: Age-Related Differences in Eye Tracking and Usability Performance: Website Usability for Older Adults. International Journal of Human-Computer Interaction, 29, 541-548.

Severin, W.J., Tankard, Jr., J.W., 2001: Communication theories: Origins, methods, and uses in the mass media (5th ed.). New York: Addison Wesley Longman. Ch. 1.

Snover, A.K., Mantua, N.J., Littell, J.S., Alexander, M.A., McClure, M.M., Nye, J., 2013: Choosing and using climate-change scenarios for ecological-impact assessments and conservation decisions. Conservation Biology, 27, 1147-1157.

Titus, S., Horsman, E., 2009: Characterizing and Improving Spatial Visualization Skills. Journal of Geoscience Education Research, 57, 242-254.

USDOI (U.S. Department of Interior), 2009: Climate change and the Department of Interior, Secretarial order 3226, amendment no. 1. USDOI, Washington, D.C.

Velasquez, J.D., 2013: Combining eye-tracking technologies with web usage mining for identifying Website Keyobjects. Engineering Applications of Artificial Intelligence, 26, 1469-1478.

Wagner, N., Hassanein, K., Head, M., 2014: The impact of age on website usability. Computers in Human Behavior, 37, 270-282.

Wilson, K.A., Heinselman, P.L., Kang, Z., 2016: Exploring Applications of Eye Tracking in Operational Meteorology Research. Bulletin of the American Meteorological Society, 97, 2019-2025.

Table 1: Research questions, the population, eye-tracking metrics, other data, and analyses used to answer the questions.

Questions	**Pop.**	**ET Metrics**	**Other Data**	**Analyses**
Gender: Are there statistically significant differences between males and females in the second phase?	Phase 2	FC, TFF, TVD	Tasks/Questions	Statistics
Gender: If so, does spatial ability play a role?	Phase 2	NA	Questionnaire, Object Slicing Test	Statistics
Gender: Are there similarities between males and females from both populations?	Phase 1, Phase 2	FC, TFF, TVD	Questionnaire, Tasks/Questions	Statistics; 1:1 plots
Expert v. Novice: How do the Phase 1 forestry professionals compare to the Phase 2 students?	All	Heat maps and gaze plots; FC, TFF, TVD	Tasks/Questions	Visual analysis; Statistics; 1:1 plots
Expert v. Novice: Are there any statistically significant differences between Phase 1 Experts and Phase 2 Novices?	Phase 1, Phase 2	Heat maps and gaze plots; FC, TFF, TVD	Tasks/Questions	Visual analysis; Statistics
Usability: Which design elements distract or confuse users?	All	Heat maps and gaze plots; FC, TFF, TVD	NA	Visual analysis; Statistics
Usability: Do the changes help users better navigate the tools and answer the questions more quickly and/or accurately?	Phase 1, Phase 2	FC, TFF, TVD	Tasks/Questions	Statistics

CHAPTER 2: Website Usability Differences between Males and Females: An Eye-tracking Evaluation of a Climate Decision Support System

Lindsay C. Maudlin, Karen S. McNeal, Heather Dinon-Aldridge, Corey Davis, Ryan Boyles, Rachel M. Atkins

[TO BE SUBMITTED FOR PUBLICATION TO THE BULLETIN OF THE AMERICAN METEOROLOGICAL SOCIETY]

Capsule

Evaluations of climate decision support systems help to identify potential barriers to successfully disseminating climate information to diverse groups of end-users.

Abstract

Decision support systems, which are collections of related information located in a central place, can be used as platforms from which climate information can be shared with decision-makers. In this study, a web-based climate decision support system (DSS) for foresters in the Southeast United States was evaluated using eye-tracking technology. The initial study design was exploratory and focused on assessing usability concerns within the website. Results showed differences between male and female forestry experts in their eye-tracking behavior and in their success with completing tasks and answering questions related to the climate information presented in the DSS. A follow-up study, using undergraduate students from a large university in the Southeast United States, aimed to determine if similar gender differences would be detected, and if so, if the cause(s) could be determined. The second evaluation, similar to the first, showed that males and females focused their attention on different aspects of the website; males focused

more on the maps depicting climate information, while females focused more on other aspects of the website (e.g., text, search bars, color bars). DSS developers should consider these gender differences when designing a web-based DSS in order to effectively support various populations of users.

Main Text

In the face of global climate change, the need to disseminate accurate climate information to decision-makers for adaptation purposes has grown; however, this is no easy feat. Consideration must be given to the climate information provided to decision-makers such that it is understandable and used appropriately (Snover et al., 2013). Further consideration should be given to the presentation of the climate information from the visualization and design of the content to the tailoring of information for the intended audience (Daron et al., 2015; Harold et al., 2016).

One way to disseminate tailored climate information to decision-makers is through a decision support system (DSS), which is defined as a collection of information located in a central place that decision-makers can access and consult before making a decision. DSSs are becoming increasingly common across many disciplines, but they are rarely, if ever, evaluated and do not always meet the needs of the decision-makers using them (Pyke et al., 2007). To ensure that this strategy is effective and that decision-makers understand the information they will ultimately use to make potentially high-impact decisions, the efficacy of these DSSs needs to be investigated through evaluations (Perry et al., 2016).

One such DSS has been developed and evaluated. As part of the Pine Integrated Network: Education, Mitigation, and Adaptation Project (PINEMAP), a USDA National Institute of Food and Agriculture project that focused primarily on loblolly pine forests in the Southeast

United States, the PINEMAP DSS was developed. The web-based PINEMAP DSS provides historical and future climate and regional productivity model data specifically for foresters in this region through a suite of tools that explore parameters such as temperature and precipitation. After initial PINEMAP DSS development, a beta testing period indicated a need for further evaluation that could address how users interacted with the PINEMAP DSS. The evaluation included two phases: Phase 1, the exploratory, initial evaluation and Phase 2, the follow-up evaluation. The initial evaluation (Phase 1) involved experts from the forestry field as these respondents most closely represented the target audience of the PINEMAP DSS. The main goal of the initial evaluation was to identify which aspects of the website impacted a user's ability to successfully navigate and use the website. Initial findings highlighted aspects of the PINEMAP DSS that served as stumbling blocks in user navigation, and they also identified key differences between males and females in their eye movements and in their success in completing tasks. Based on the initial findings from Phase 1, the PINEMAP DSS was modified to address the identified stumbling blocks. The follow-up evaluation (Phase 2) involved undergraduate students and aimed to determine if the modifications made the PINEMAP DSS easier to use and if the differences between males and females found in Phase 1 persisted after implementing these changes.

Study design and populations

All study participants were given five minutes to freely explore the PINEMAP DSS website. The free exploration period allowed participants to become familiar with the website and gave them an opportunity to read about the PINEMAP DSS or to begin using the tools. At the end of the five minutes, participants were automatically advanced to the second portion of the study which measured the usability of the PINEMAP DSS and involved completing one task and answering two multiple choice questions for each tool in the evaluation.

Three tools were evaluated: Extreme Minimum Temperature; Summer Precipitation; and Seedling Market (later renamed to Cold-Tolerant Markets for Nurseries in Phase 2). Altogether, there were three tasks and six multiple choice questions. The tasks were designed to utilize certain features of the PINEMAP DSS, one at a time, in order to measure the usability of individual website features. Additionally, each task built on the knowledge gained from the previous tasks. For example, the first task asked the users to input latitude and longitude coordinates, and the last task asked the users to enter the latitude and longitude coordinates, select a time period, and find a drop-down menu to turn-on a specific map layer.

Participants' eye movements were recorded throughout their interactions with the PINEMAP DSS through eye tracking, a non-invasive technology that determines the location and duration of users' visual fixations. User demographics (age, gender, and education level) were also collected to determine if certain characteristics influenced a user's success in completing the tasks and answering the related questions correctly. Additionally, participants were asked open-ended questions about their experiences using climate data and climate information websites.

The participant group from Phase 1 included 30 forestry students and professionals (18 males and 12 females; 21 to 65 years old), and their highest degrees, either completed or in progress, ranged from undergraduate to doctoral degrees (6 undergraduate degrees, 17 master's degrees, and 6 doctoral degrees). They were recruited at a forestry conference in Durham, North Carolina, and were compensated afterward with their choice of a pre-packaged candy bar or a rain gauge. The participant group from Phase 2 included 12 undergraduate students (4 males and

8 females; 18 to 21 years old) in STEM and non-STEM majors, and their grade levels ranged from freshmen to juniors (2 freshmen, 8 sophomores, and 2 juniors). They were recruited from introductory physical science classes at a large research institution in the Southeast United States and were compensated afterward with a pre-packaged candy bar.

In addition to the study protocol already described, participants in Phase 2 were asked about their experiences with activities linked to developing spatial skills and/or those which develop map reading skills such as scouts, building blocks, playing video games, and using maps for navigation (Casey et al., 2008; Jirout and Newcombe, 2015; Gold et al., 2018). They also completed a block slicing test (Titus and Horsman, 2009; Ormand et al., 2014) to measure one aspect of spatial ability, visual penetrative ability. These activities were added to the study design to help determine if differences in spatial ability and spatial skills could explain any potential differences between the eye movements and successful completion of tasks of males and females.

Both phases of this study were carried out under approval through the Institutional Review Board for the Protection of Human Subjects. To ensure validity, the tasks and questions were written with the help of experts in the content area, and the complete study was reviewed by multiple graduate students. Additionally, the authors are aware of the following limitations: the two study populations were not identical; changes were made to the PINEMAP DSS website between the two evaluation phases, so the website was not identical between Phase 1 and Phase 2; and, due to limited time with participants, only one instrument was used to measure spatial ability. Further, the study sample size of males and females was limited due to the emphasis on the eye-tracking applications (typical sample sizes are less than 30 subjects in eye-tracking research) and may or may not be generalizable to the targeted study population or broader public audiences. Future work is suggested in order address the limitations in this current study.

Differences between males and females

Guided by the existing literature on gender differences in website usability studies (Simon, 2001; Cyr and Bonanni, 2005; Moss et al., 2006; Djamasbi et al., 2007), a major focus of the initial PINEMAP DSS eye-tracking evaluation and subsequent follow-up study was on identifying any differences between males and females. Eye tracking revealed qualitative differences in the visual fixation patterns of Phase 1 males and females throughout many of the PINEMAP DSS features, but the most obvious differences were identified in the Seedling Markets tool (Figure 1), the third of three tools used in the evaluation. For this particular tool, participants were tasked with entering a specific latitude and longitude and toggling on a feature by clicking a button located above the center map. One of the questions associated with this task required that the participants look at the city located at the specified latitude and longitude and compare the information for the city between the three maps (the center map and the two side maps). The male visual fixation pattern showed a major hotspot (50+ visual fixations) over the city location in the center map and two smaller hotspots over the boxes where latitude and longitude coordinates were entered, indicating that the majority of male visual fixations were focused on the center map (Figure 1). The female visual fixation pattern showed a broader range of hotspot locations throughout the Seedling Markets tool (Figure 1). Female visual fixations focused more heavily on the non-map aspects of the website such as various features that could be toggled off and on, the latitude and longitude boxes, drop-down menus, and the map layer options menu box. While some attention is given to the data, the majority is not.

In Phase 2, a few modifications were made to the Seedling Markets tool, but the task and related questions were unchanged. The tool was renamed to Cold-Tolerant Markets for Nurseries, and the map layer options menu box was left open by default (it was minimized by default in Phase 1). The visual fixation patterns of males and females were more similar to each other than those from Phase 1, but there were some key differences (Figure 2). The Phase 2 males and females had similar visual fixation patterns for the content located above the maps; however, females gave more attention to the center map and failed to give much attention to the side maps as compared to the males, meaning that females were not looking at all of the required data before answering the related question. Additionally, the most attention was given to the map layer options menu box.

In general, males in both phases fixate more often on the map aspects that will help to answer the related questions at the end of the task (e.g., 14 times v. 6 times, on average, for the city location) while females fixate more often on other aspects of the DSS (e.g., 28 times v. 19 times, on average, for the map layer options menu box). A notable difference between males and females in both phases is how they used their free exploration time at the start of the study. The free exploration period was an opportunity for participants to become familiar with the PINEMAP DSS website and tools. With their free exploration time, males fixated more on the data (the maps themselves) than did females, and in some cases, the females did not fixate anywhere within a given AOI (Figure 3). Additionally, females fixated more on the text on the home page as they read through the material than did males (e.g., 169 times v. 148 times, on average).

During the tasks and related questions portion of the study, males fixated more than females on the location of the city of interest within the maps and the surrounding data; and females fixated more on the latitude and longitude boxes and other DSS features outside the maps than did males (Figures 1 and 4). By giving more attention to the city location within the maps, males are likely taking in more information that will ultimately help them answer the related questions. By giving more attention to the website aspects above the maps, females are likely becoming more familiar with the various features of the website but not gaining insight into the data that will help answer the related questions. When combining these two major differences between males and females, how they use their free exploration time and where they focus their attention most during the tasks and while answering the related questions, differences in their success of answering the questions correctly can be expected.

Out of the six questions asked in the tasks and related questions portion of the study, males answered more questions correctly (Figure 5), on average (4.74 v. 4.21). Out of all subgroups, Phase 2 males answered the most questions correctly, on average (5.25), and Phase 2 females answered the least questions correctly, on average (3.88), resulting in a large effect size difference between male and female performance (Cohen's d = 0.81). The three lowest performers from Phase 2 (those with total scores less than or equal to 3 out of 6) in this study were all females (Figure 5). Additionally, they also had the least experience with any activities related to developing spatial skills. Out of the twelve spatial-related activities inquired about in this study, males had experience with an average of 8.75, females had experience with an average of 6.38, and the three lowest performers had experience with an average of 3.33. The lowest performers also had the lowest scores on the spatial ability test (3.33) compared to males (6.25) and females (4.66). While experience with these types of activities and the spatial ability test used in this study do not directly predict performance with the tasks and related questions portion of the study, the similarities between the three lowest performers suggest these are factors that can influence performance.

Implications for future climate decision support systems

The findings from this study highlight the influence of end-user demographics and experience level on the development of DSS products. For example, if we want to ensure females focus on the data found within the maps, extra effort should be given to ensure that these users are visually directed to that information (e.g., use an arrow or a pulsing banner to draw attention to the map). Similarly, if we want to ensure that males read about a tool before using it, extra effort should be given to drawing their attention to the relevant text (e.g., include the necessary information on the same page as the tool itself and highlight it with a pulsing banner).

Additional research and effort are needed to further address how gender might impact a user's success in navigating and interpreting information from a climate DSS, and other user demographics and experiences should also be considered. To ensure the development of a usable and easily understood climate DSS for decision-makers, the target audience should be clearly identified. In addition, evaluation should be included from the beginning of the design process (Perry et al., 2016) and should draw on the design-based research and evaluation frameworks (McNeal et al., 2014) used in this study.

Sidebar: Eye tracking

Eye tracking is a non-invasive, harmless technology that allows researchers to determine where, when, and for how long a user is looking at specific elements on a computer screen. It has been used for more than 100 years (Jacob and Karn, 2003) in various applications, including website usability studies (Holmqvist et al., 2011; Drost et al., 2015; Wilson et al., 2016; Bojko, 2006; Loyola et al., 2015).

The eye-tracking hardware used to evaluate the PINEMAP DSS utilized two different types of instruments. The Phase 1 set-up consisted of two Tobii X2-60 (60 gaze points per second) eye trackers used with portable laptop computers (at a forestry conference with forestry experts). The Phase 2 set-up consisted of one Tobii TX-300 (300 gaze points per second) eye tracker used with a stationary desktop computer (on campus with undergraduate students). Before eye movements were recorded, each participant underwent a calibration process and was seated such that the eyes were ~65 cm from the eye tracker, all to ensure accurate and consistent eye-tracking data. After the eye-tracking data were collected, the Tobii software provided both qualitative and quantitative outputs for further analysis.

The qualitative outputs include heat maps, which indicate where and how many visual fixations occur. These products are suitable for visual analysis and general descriptions of user eye movements and behaviors but not for statistical analysis purposes.

The quantitative outputs are calculated based on areas of interest (AOIs), regions within the PINEMAP DSS that are defined by the researcher. For example, one AOI is the region in each of the PINEMAP DSS tools where the user would enter in the latitude and longitude coordinates for a specific location. Three quantitative metrics are calculated specific to each AOI: the fixation count (FC), the total number of times that a user fixates anywhere within the AOI; the time to first fixation (TFF), the total number of seconds between the time at which the webpage loads and the time at which the user fixates for the first time within the AOI; and the total visit duration (TVD), the total number of seconds over all visits to the AOI that the user fixates within the AOI. These metrics are exported from Tobii for additional analyses. For quality control purposes, the eye-tracking data (i.e., FC, TFF, and TVD) from participants with low calibration rates (<60%) are not included in the statistical analyses throughout this study.

Sidebar: Gender differences

The literature on gender differences within website design and usability studies is limited. Moreover, existing studies are heavily focused on e-commerce and online shopping applications and the aesthetics of these types of websites; however, these differences can shed some light on design preferences between the genders that can still be relevant to the PINEMAP DSS evaluation in this study. Cyr and Bonanni (2005) show that women are consistently less pleased with a website than men using a Likert scale rating system. In addition, this study indicates that more men than women approve of the website's organization and layout and are able to easily navigate the website. Other studies have reported gender differences in design preferences including shapes and colors and in opinions of what is eye-catching and attractive (Moss et al., 2006; Djamasbi et al., 2007). One possible explanation for these gender differences is the manner in which males and females process information. Males tend to be selective with the information that they attempt to comprehend while females tend to attempt to comprehend all of the available information, so females are more likely to be less satisfied with a website if it does not include adequate information or quickly draw their attention to it (Simon, 2001).

If users are expected to find information and be able to recall and apply that information, then a second factor, search behavior, must be considered as this can impact learning gains (Roy et al., 2003; Roy and Chi, 2003). For example, important design criteria for males include minimal action, meaning that less clicking or typing to navigate the website, and flexibility, meaning tasks can be completed in different ways. In contrast, females prefer learnability, which entails having a similar layout and navigation process throughout the website, and user guidance, which entails feedback and hints (Lin and Hsieh, 2016).

The male-dominated IT industry could potentially be a stumbling block for women using websites mainly designed by males (Moss et al., 2006), in part because previous research has shown that gender preferences of websites and their elements tend to align best when the same gender designed the site (Moss and Gunn, 2007). Boys have been shown to have higher aesthetic preferences for existing children's websites than girls, in part due to visual complexity since boys preferred higher design complexity than girls (Wang, 2014). Less studied is the affective aspect of website design and how it might influence website usability. A user's affective response to a new website has been shown to influence his or her perceived usefulness and ease of use (Zhang and Li, 2005). Statistically significant gender differences justify customizing a website for its audience (Simon, 2001), and if a website is not customized for female viewers or designed with them in mind, females may experience frustration and anxiety and consequently be less satisfied with the website (Lin and Hsieh, 2016). Additional research on gender differences in website design and usability is needed to better inform website developers to be more female-friendly.

References

Casey, B. M., N. Andrews, H. Schindler, J. E. Kersh, A. Samper, and J. Copley, 2008: The Development of Spatial Skills Through Interventions Involving Block Building Activities, Cognition and Instruction, 26, 269-309, https://doi.org/10.1080/07370000802177177.

Cyr, D., and C. Bonanni, 2005: Gender and website design in e-business. International J. of Electronic Business, 3, 565-582.

Daron, J. D., S. Lorenz, P. Wolski, R. C. Blamey, and C. Jack, 2015: Interpreting climate data visualizations to inform adaptation decisions. Climate Risk Management, 10, 17-26.

Djamasbi, S., T. Tullis, J. Hsu, E. Mazuera, and K. Osberg, 2007: Gender Preferences in Web Design: Usability Testing through Eye Tracking. Proc. of the Thirteenth Americas Conference on Information Systems, Keystone, Colorado, 1-8.

Gold, A. U., P. M. Pendergast, C. J. Orman, D. A. Budd, J. A. Stempien, K. J. Mueller, K. A. Kravitz, 2018: Spatial skills in undergraduate students—Influence of gender, motivation, academic training, and childhood play. Geosphere, 14, 668-683.

Harold, J., I. Lorenzoni, T. F. Shipley, and K. R. Coventry, 2016: Cognitive and psychological science insights to improve climate change data visualization. Nature Climate Change, 6, 1080-1089.

Jirout, J. J., and N. S. Newcombe, 2015: Building Blocks for Developing Spatial Skills: Evidence From a Large, Representative U.S. Sample. Psychological Science, 26, 302-310, https://doi.org/10.1177/0956797614563338.

Lin, C. J., and T. –L. Hsieh, 2016: Exploring the design criteria of website interfaces for gender. International Journal of Industrial Ergonomics, 53, 306-311.

McNeal, K. S., J. C. Libarkin, T. S. Ledley, E. Bardar, N. Haddad, K. Ellins, and S. Dutta, 2014: The Role of Research in Online Curriculum Development: The Case of EarthLabs Climate Change and Earth System Modules. J. of Geoscience Education, 62, 560-577.

Moss, G. A., R. Gunn, and J. Heller, 2006: Some Men Like It Black, Some Women Like It Pink: Consumer Implications of Differences in Male and Female Website Design. J. of Consumer Behaviour, 5, 328-341.

Moss, G. A., and R. Gunn, 2007: Gender Differences in Website Design: Implications for Education. Systemics. Cybernetics and Informatics, 5, 38-43.

Ormand, C. J., C. Manduca, T. F. Shipley, B. Tikoff, C. L. Harwood, K. Atit, A. P. Boone, 2014: Evaluating Geoscience Students' Spatial Thinking Skills in a Multi-Institutional Classroom Study. J. of Geoscience Education, 62, 146-154.

Perry, S. C., and Coauthors, 2016: Get your science used—Six guidelines to improve your products: U.S. Geological Survey Circular, 1419, https://doi.org/10.3133/cir1419.

Pyke, C. R., B. G. Bierwagen, J. Furlow, J. Gamble, T. Johnson, S. Julius, and J. West, 2007: A decision inventory approach for improving decision support for climate change impact assessment and adaptation. Environmental Science & Policy, 10, 610-621.

Roy, M., and M. T. H. Chi, 2003: Gender Differences in Patterns of Searching the Web. J. of Educational Computing Research, 29, 335-348.

Roy, M., R.Taylor, and M. T. H. Chi, 2003: Searching for Information On-line and Off-line: Gender Differences Among Middle School Students. J. of Education Computing Research, 29, 229-252.

Simon, S. J., 2001: The Impact of Culture and Gender on Web Sites: An Empirical Study. The DATA BASE for Advances in Information Systems, 32, 18-37.

Snover, A. K., N. J. Mantua, J. S. Littell, M. A. Alexander, M. M. McClure, and J. Nye, 2013: Choosing and using climate-change scenarios for ecological-impact assessments and conservation decisions. Conservation Biology, 27, 1147-1157.

Titus, S., and Horsman, E., 2009: Characterizing and Improving Spatial Visualization Skills. J. of Geoscience Education, 57, 242-254.

Wang, H.-F., 2014: Picture Perfect: Girls' and boys' preferences towards visual complexity in children's websites. Computers in Human Behavior, 31, 551-557.

Zhang, P., and N. Li, 2005: The Importance of Affective Quality. Communications of the ACM, 48, 105-108.

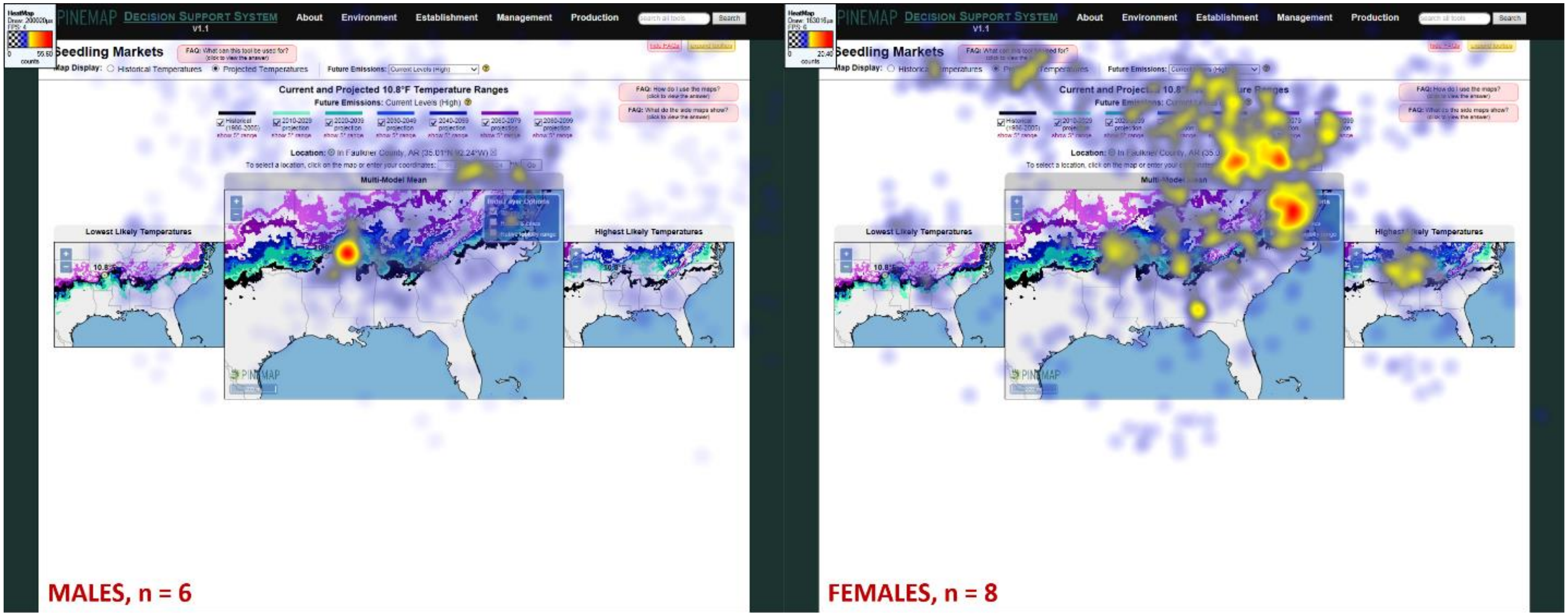


Figure 1. Heatmaps for a subset of the males (left) and females (right) for Phase 1 Seedling Markets Tool (renamed to Cold-Tolerant Markets for Nurseries in Phase 2). Red denotes high visual fixation counts, and blue denotes low visual fixation counts. Males fixate more on the data within the center map while females fixate more on the other website aspects.

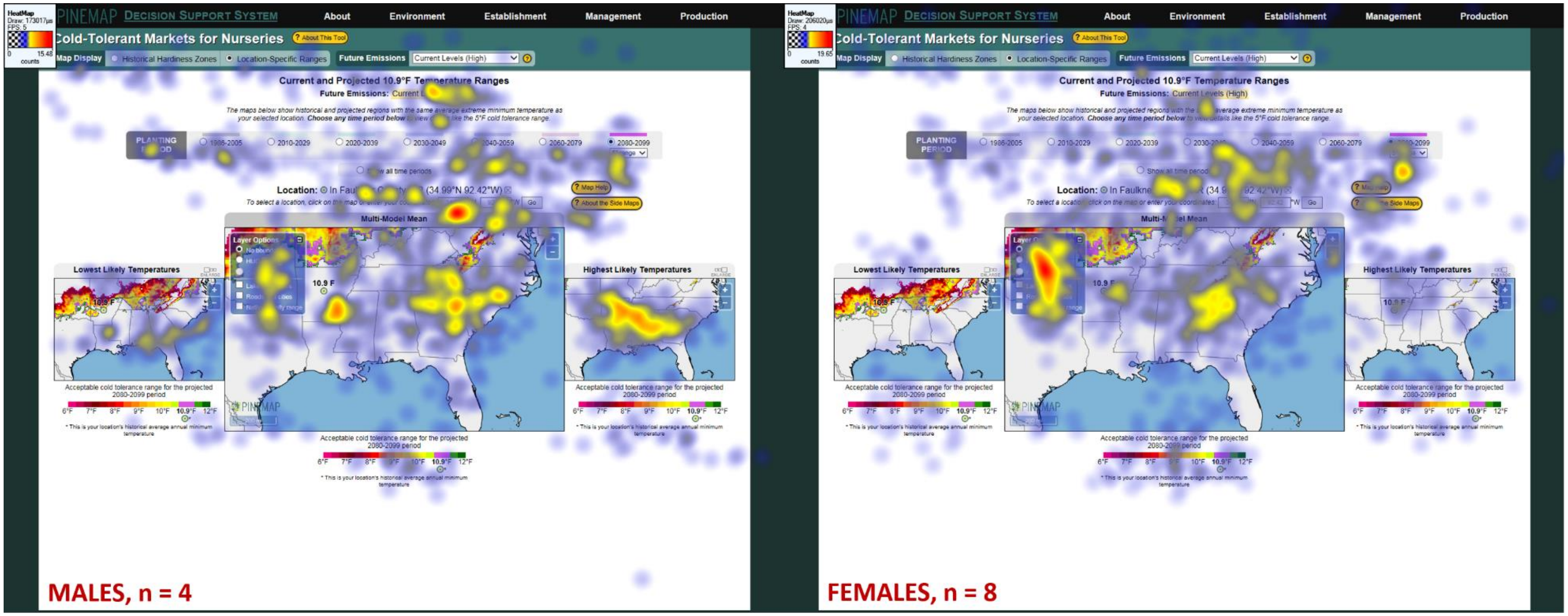


Figure 2. Heatmaps for males (left) and females (right) for the Phase 2 Cold-Tolerant Markets for Nurseries (previously named Seedling Markets Tool in Phase 1). Red denotes high visual fixation counts, and blue denotes low visual fixation counts. Males and females have similar visual fixation patterns for the content located above the center map, and both look at the data within the center map. Males give more attention to the map on the right, while females neglect the side maps almost entirely.

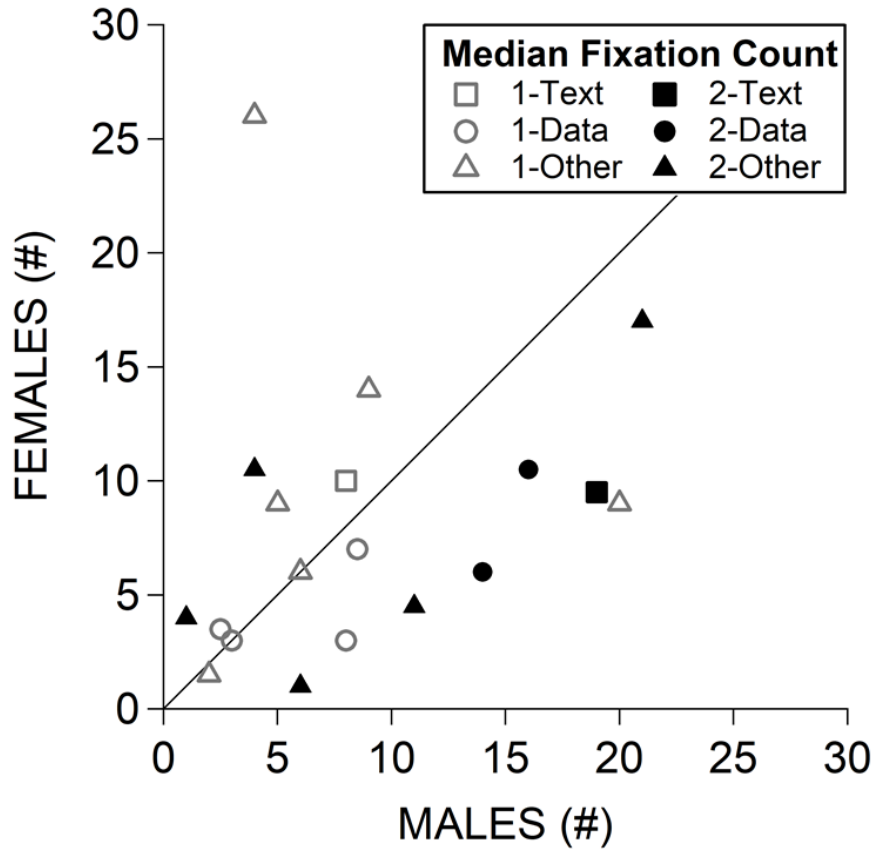


Figure 3. A comparison of males and females and how and where they focus their attention during the five minutes of free exploration at the start of the study. Markers represent the median value of visual fixation counts of males compared to females for individual areas of interest (AOIs). Those above the diagonal line mean that females had a higher median visual fixation count than males, and those below the diagonal line mean that males had a higher median visual fixation count than females. Markers are also differentiated between Phase 1 (denoted with “1-“) and Phase 2 (denoted with “2-“) and by type of AOI, where “Text” refers to AOIs filled with blocks of text, “Data” refers to AOIs located within the maps, and “Other” refers to AOIs that include color bars, latitude and longitude boxes, and various other buttons. Two AOIs from the “2-Data” category are missing from this figure because only the males had visual fixation counts for these particular AOIs.

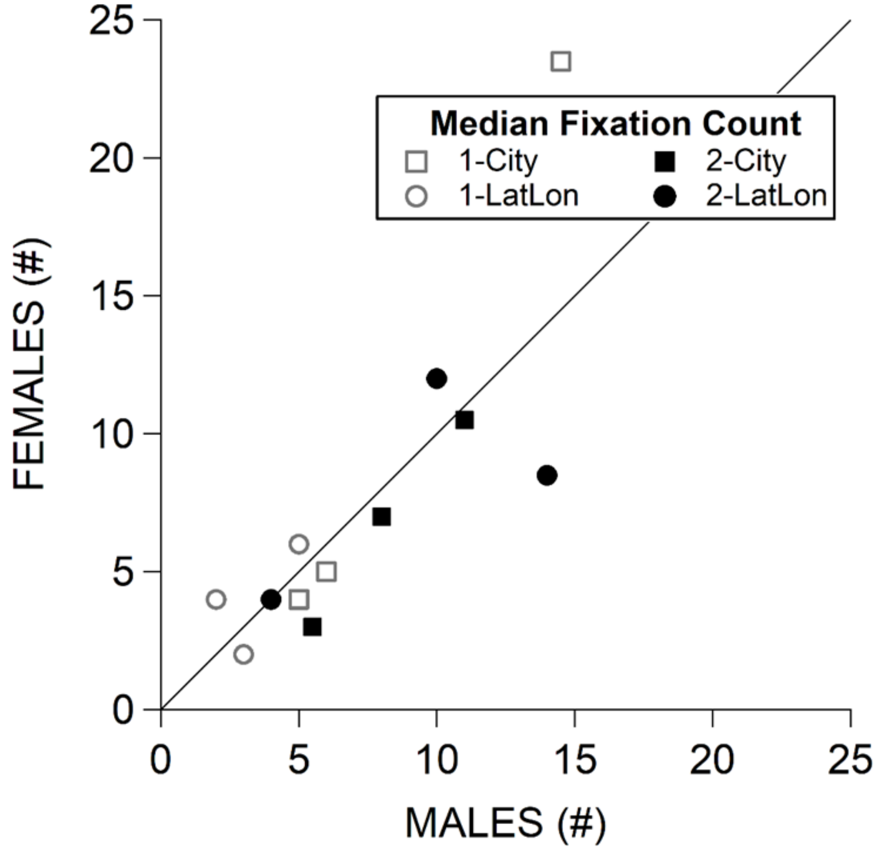


Figure 4. A comparison of males and females and how and where they focus their attention during tasks and while answering the related questions. Markers represent the median value of fixation counts of males compared to females for individual areas of interest (AOIs). Those above the diagonal line mean that females had a higher median fixation count than males, and those below the diagonal line mean that males had a higher median fixation count than females. Markers are also differentiated between Phase 1 (denoted with “1-“) and Phase 2 (denoted with “2-“) and by type of AOI, where “City” refers to the location of the specified city for each of the three tools, an indication that participants were looking at the map data, and “LatLon” refers to the latitude and longitude boxes above the maps, an AOI that is indicative of attention being given to the content above the maps.

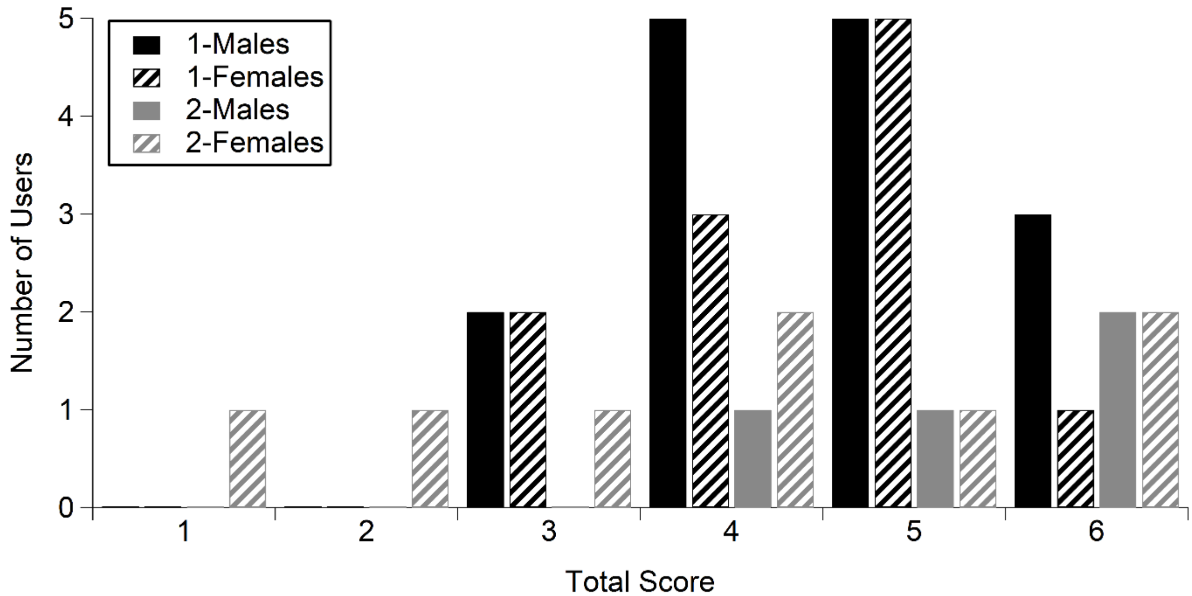


Figure 5. Total score out of six possible for males and females in Phase 1 (denoted with “1-“) and Phase 2 (denoted with “2-“).

CHAPTER 3: The Role of Expertise on the Usability of a Climate Decision Support System

Lindsay C. Maudlin, Karen S. McNeal, Heather Dinon-Aldridge, Corey Davis, Ryan Boyles,

Rachel M. Atkins

[TO BE SUBMITTED FOR PUBLICATION TO WEATHER, CLIMATE, & SOCIETY]

Abstract

A web-based climate decision support system (DSS) was evaluated using eye tracking. Users were asked to complete tasks and answer related multiple choice questions with interacting with the DSS. The users represented a spectrum of expertise from undergraduate students to professionals with doctorate degrees and a broad range of ages. In general, the experts were outperformed by some novices, and previous experience using climate data did not guarantee success. Furthermore, novices had relatively shorter time to first fixation on DSS elements relevant to the tasks and questions, but they had relatively longer total visit durations. The poor performance and the longer time to first fixation of experts in this study are likely due to their relatively older ages compared to other users. This study and its findings suggest that age is a confounding factor in web-based climate DSS usability.

Introduction

In order for communities to successfully adapt to global climate change, decision-makers must be given accurate and relevant climate information in an understandable format that draws clear connections between the projected climate impacts and their communities. For this reason, effort has been put into learning how to successfully communicate science with the general public and policymakers such that they understand and know how to apply the information they are given (Romsdahl, 2011; Perry et al., 2016). Scientists can draw on effective strategies from the communication field such as interacting with and framing the message for the intended audience (Nisbet, 2009; Nisbet and Scheufele, 2009; Scannell and Gifford, 2013).

Framing a climate information message for a specific group of decision-makers can be done through a decision support system (DSS). A DSS combines all relevant information into a central location that is easily accessible by decision-makers. DSSs are commonly used across many fields but less commonly evaluated (Pyke et al., 2007). Without evaluations, the efficacy of DSSs cannot be known, and scientists are unable to determine if the products they develop for a specific group of end-users are usable, easily understood, or provide the relevant climate information for their intended audience. The increased use of web-based DSSs introduce the need to consider website usability.

Website usability is commonly studied across many disciplines, and these studies have identified various user characteristics such as age and gender that influence website usability. Older users are slower in website navigation and are oftentimes less accurate as they complete tasks (Romano Bergstrom et al., 2013), and the loss in cognitive skills with age has been found to decrease spatial ability, mental model accuracy, and usability (Wagner et al., 2014; Grahame et al., 2004).

An understanding of how user characteristics influence how they interpret visualizations is also critical. As visualizations have generally improved in their usage and usability, they are typically designed in a single format for all users that does not consider diverse user characteristics (Toker et al., 2013). One such example is user expertise. Expertise is a good predictor of performance and impacts visualization effectiveness (Toker et al., 2012), and for this reason, domain-based experts and novices have been compared across many different disciplines

(Barfield, 1986; Simmons and Lunetta, 1993; Chen et al., 2006; Hmelo-Silver et al., 2007; Jarodzka et al., 2010; Gegenfurtner et al., 2011; Kastens et al., 2016). Experts are usually scientists or college-level instructors with formal training and experience in a particular domain, while novices are typically students with little or no training in the same domain (Simmons and Lunetta, 1993). Furthermore, what experts know within a specific domain is not just quantitatively different but also qualitatively different (LaFrance, 1989). Experts have been shown to remember more lines of code than novices due to their ability to correctly group and connect the information they are looking at rather than simply having a greater mental capacity (Barfield, 1986). When looking at complex and dynamic visualizations, experts tend to focus their attention on the more relevant information surrounded by irrelevant information than do novices, demonstrating perceptual/attentional skills (Jarodzka et al., 2010). Experts are also often able to complete tasks faster than novices (Jarodzka et al., 2010).

Many expert-novice studies utilize eye tracking (Jarodzka et al., 2010; Gegenfurtner et al., 2011; Ooms et al., 2012; Kastens et al., 2016). Through this technology, experts have been shown to fixate, focus their gaze and attention on a certain element, more on relevant data for answering specific questions and completing tasks than do novices (Jarodzka et al., 2010; Gegenfurtner et al., 2011; Atkins and McNeal, in review). Furthermore, experts have been shown to have shorter total fixation durations and shorter times to first fixation on relevant information (Gegenfurtner et al., 2011). Experts have been shown to have significantly shorter fixations and more fixations per second while completing a visual search through a series of maps, demonstrating their ability to interpret a larger part of the map in the same amount of time than novices (Ooms et al., 2012).

Given the nature of studies that examine differences between experts and novices, sample sizes are often quite small and the findings lead to more questions, indicating the need for additional studies (Simmons and Lunetta, 1993). This study addresses several gaps in the literature. First, it employs eye-tracking data to evaluate the usability and understandability of a climate DSS. Second, it explores website usability with users of varying levels of expertise. Third, it quantifies the differences between eye movements of experts and novices as they interact with complex climate information. The key questions this study aims to answer are: 1) Do those with higher levels of education and more experience using climate data out-perform those with less education and experience?; 2) Are there differences in the eye-tracking behavior between experts and novices?; and 3) If differences are present, are they significant?

Methods

Study design

A climate DSS is evaluated in two phases. The first phase (Phase 1) was exploratory and focused on identifying usability problems within the climate DSS. Preliminary results from Phase 1 prompted DSS updates and changes as well as a follow-up evaluation (Phase 2). Phase 2 focused on determining whether DSS updates and changes improved the overall usability of the DSS.

All users were given five minutes to freely explore the DSS. During this time, users could read about the DSS, explore the various tools, or any combination of those two activities. At the end of the free exploration period, users began a series of three tasks, one for each DSS tool evaluated, and two multiple choice questions per task (Supplemental Material). The tasks were developed with the intent of measuring the usability of the DSS. They became more difficult with each task as they built on the prior knowledge gained from the previous task. The multiple choice questions assessed whether participants understood what they were looking at. Eye tracking was utilized throughout this portion of the study.

After interacting with the DSS, users were asked to provide demographic data (age, gender, and education level). Demographic data were collected to determine if certain characteristics influenced a user's ability to complete the tasks and answer the related questions correctly. Additionally, participants were asked open-ended questions about their experiences using climate data and climate information websites to determine how familiar users already were with climate data.

Study populations

Participants in Phase 1 were forestry professionals and recruited at a forestry conference in Durham, North Carolina. There were 30 participants in total, 18 males and 12 females, and their ages ranged from 21 to 65. Their highest degrees (completed or in progress) included 6 undergraduate degrees, 17 master's degrees, and 6 doctoral degrees in forestry or similar fields. Compensation was their choice of a pre-packaged candy bar or a rain gauge.

Participants in Phase 2 were undergraduate students (STEM and non-STEM majors) recruited from introductory physical sciences classes at a large research university in the Southeast United States. There were 12 participants in total, 4 males and 8 females, and their ages range from 18 to 21. Their current grade levels included 2 freshmen, 8 sophomores, and 2 juniors. Compensation was a pre-packaged candy bar.

Both phases of this study were carried out under approval through the Institutional Review Board for the Protection of Human Subjects.

PINEMAP DSS

The Pine Integrated Network: Education, Mitigation, and Adaptation Project (PINEMAP) led to the development of a DSS, PINEMAP DSS, which provides tailored climate and regional productivity model data for foresters in the Southeast United States (Davis et al., in review). PINEMAP DSS is presented in a web-based format and contains several tools that allow users to explore how the climate will vary across the region and over future time periods. Three tools were chosen to be used in the evaluation: Extreme Minimum Temperature (EMT), which shows projected changes in the occurrence of minimum temperatures below different thresholds; Summer Precipitation (SP), which shows projected changes in average summertime precipitation; and Seedling Markets (SM), later renamed the Cold-Tolerant Markets for Nurseries (CTMN), which shows northerly regions where seedlings could be deployed in the future without an appreciable change in cold risk.

Eye tracking

The PINEMAP DSS evaluations utilized different eye-tracking set-ups for the two phases. Two Tobii X2-60 (60 gaze points per second) eye trackers were used with portable laptop computers during Phase 1 to collect data remotely, and one Tobii TX-300 (300 gaze points per second) eye tracker was used with a stationary desktop computer during Phase 2 to collect data within a lab on campus. After the eye-tracking data were collected, the Tobii Studio software provided both qualitative and quantitative outputs for analyses. Aside from the two different frequencies (hertz) of the Tobii systems, all other parameters were identical during eye-tracking and follow-on analysis.

The qualitative outputs consist of heat maps and gaze plots. Heat maps indicate where individuals or groups of people fixate and how many fixations occur. Gaze plots indicate when and where users fixate and the paths their eyes take between fixations. These products are suitable for visual analysis and general descriptions of user eye movements and behaviors but not for statistical analysis purposes.

The quantitative Tobii Studio outputs are calculated based on areas of interest (AOIs), regions within the PINEMAP DSS that are defined by the researcher. For example, one AOI is the region in each of the tools where the user would enter the latitude and longitude coordinates for a specific location provided in a task. Once the AOIs are defined, three quantitative metrics are then calculated specific to each AOI: the fixation count (FC), which is the total number of times that the user fixates anywhere within the AOI; the time to first fixation (TFF), which is the total number of seconds between the time at which the webpage loads and the time at which the user fixates for the first time within the AOI; and the total visit duration (TVD), which is the total number of seconds over all visits to the AOI that the user fixates within the AOI. These metrics are then exported from Tobii Studio to be used in statistical analyses.

Before eye movements were recorded, all participants were seated such that their eyes were ~65 cm from the eye tracker, and each underwent a calibration process to ensure accurate and consistent eye-tracking data. Data from participants with poor calibrations or low sample rates (<60%) are not included.

Additional data

Questionnaires (Supplemental Material) were given at the end of the eye tracking portion of the study. They aided in the collection of demographic information (e.g., age, gender, education level) and allowed participants to describe their experience with using historical climate data and climate projections through open-ended responses. The open-ended responses led to the formation of two groups: those who had experience with climate data and those who had not.

The six total multiple choice questions (Supplemental Material), two for each of the three tasks completed in the evaluation, were scored based on the correctness of the response. For each correct response, participants were given one point. The highest score possible was a total score of six points.

Limitations

Due to the inherent differences between the forestry professionals from Phase 1 and the undergraduate students in introductory physical science courses from Phase 2, the populations are not identical. Additionally, updates and changes were made to the PINEMAP DSS website between the two phases. These population differences and DSS updates and changes could influence the results that compare metrics between participants in the two phases.

Eye-tracking studies often have small sample sizes (<40) (Jacob and Karn, 2003), thus findings from individual eye-tracking studies can be influenced by sampling error or disagree with previous findings (Gegenfurtner et al., 2011).

Expert-novice differences can be influenced by task difficulty, the time spent working through a task, and the degree to which users control their own process of working through a task (Gegenfurtner et al., 2011). In this study, the tasks increase in difficulty from the first to the third, users had an unlimited amount of time to work through each task, and users had complete control over the process they could take to work through the tasks and answer the related questions. This study design will have inherently different results than one with time limits for completing tasks.

Validity

The tasks and questions specific to the PINEMAP DSS website were written collaboratively with the developers of the PINEMAP DSS and two geoscience education researchers. Once the tasks and questions were incorporated into the eye-tracking study design, the entire study was reviewed with the help of four geoscience graduate students for content validity.

Results and Discussion

Of the 30 participants in Phase 1 and 12 participants in Phase 2, data from 26 and 12 participants, respectively, are included in the following analyses (Table 1). One participant from Phase 1 did not include his/her age and is thus not included in the analyses that compare eye-tracking metrics and total scores across ages. The expert to novice spectrum ranges from the professionals with doctorate degrees in Phase 1 to the undergraduate students in Phase 2.

In terms of answering the six multiple choice questions correctly, the highest educated experts (those with doctorate degrees) performed worse than the other groups (an average total score of 4.00 compared to 4.77 for those with Master's degrees, 4.50 for professionals with Bachelor's degrees, and 4.33 for undergraduate students) (Table 1). To further investigate this phenomenon, a comparison of participants' age and total score is made (Figure 1). Excluding the undergraduate students' scores on the far left, the trend shows a general decrease in total score with age. Interestingly, those with doctorate degrees also happen to be some of the oldest participants in this study. Given this is a web-based DSS and older adults are known to struggle with website usability, the decrease in total score with age is not surprising; however, some novices outperforming some experts is unexpected.

The general distribution of users' total scores is left-skewed. Examining the distributions by degree-type (Figure 2), the patterns become more complicated. Those with doctorate degrees have total scores clustered around 4, those with Master's degrees have total scores that are left-skewed and all greater than or equal to 3, professionals with Bachelor's degrees have total scores that are also left-skewed and range from 3 to 5, and undergraduate students have total scores that are highly dispersed and span the entire range of total possible scores (1 to 6). When previous experience using climate data is considered (Table 1; Figure 2), those with previous experience using climate data are not necessarily the same participants who answer the most questions correctly. In fact, five of the seven lowest performers reported previous experience with climate data.

For the following comparisons, users with Master's degrees are used as the comparison group for several reasons: 1) they had the highest average total score of all groups (4.77); 2) the majority of them indicated they had previous experience using climate data (10 out of 13); and 3) their average age of 35 falls between the average ages of the other groups (compared to 48, 31, and 20 for those with doctorate degrees, professionals with Bachelor's degrees, and undergraduate students, respectively) (Table 1).

For most AOIs in this study, median TFF and TVD values show different trends for each degree-type, especially between the undergraduate students and those with doctorate degrees. Consider the AOIs in the SP Tool and the SM/CTMN Tool (Figures 3 and 5, respectively). Across most AOIs, the undergraduate students tend to have shorter median TFF values than professionals with Master's degrees, and sometimes even the shortest median TFF values compared to all groups (Figures 4 and 6, respectively). Conversely, the undergraduate students tend to have the longer median TVD values than professionals with Master's degrees (Figures 4 and 6, respectively). Similar trends appear when the average TFF and TVD values are compared between experts (forestry professionals) and novices (undergraduate students) in this study, in some cases demonstrating medium to large effect sizes (Table 2).

These differences demonstrate that the undergraduate students easily find the various elements of the PINEMAP DSS tools that are relevant for completing the tasks and answering the related questions, but they fixate on them longer than most other users for one of two possible reasons: 1) they have trouble comprehending the content or 2) they find the content interesting. Drawing on the known limitations of this study, specifically that comparisons between users in Phase 1 and Phase 2 are potentially biased due to the updates and changes made to the PINEMAP DSS between the two phases, the relatively lower TFF values for the undergraduate students might be influenced by overall usability improvements between the PINEMAP DSS version the professionals used (Phase 1) and the one used by the students (Phase 2). Additionally, their relatively lower average age likely plays an important role in their ability to quickly scan the PINEMAP DSS website. The relatively longer TVD values for the undergraduate students are consistent with previous research that shows that experts tend to have shorter TVD values because they require less effort to interpret the visualizations they interact with within their given domain expertise (Gegenfurtner et al., 2011). The difference in TFF values between experts in this study and other studies, namely that TFF values in this study are longer for experts than for novices, is likely due to the experts being relatively older than the novices.

Conclusion

This study examines the differences in eye-tracking behavior and overall task and question performance between experts and novices as they interact with a climate DSS. The experts in this study were defined as those with higher levels of education and more previous experience using climate data. These experts were also on the older end of the span of ages represented by study participants. This study demonstrates that the usability of a web-based climate DSS can be confounded by user expertise as well as user age. Future evaluations of web-based DSSs should consider that influences from user age on overall website usability might supersede any anticipated differences between experts and novices. For this reason, more evaluations are needed.

Recommendations from these findings and others can be made for future web-based DSS development for diverse user groups. One way to decrease differences between experts and novices is to draw novice attention to the most relevant information (Jarodzka et al., 2010). This can be easily tackled in website design by including arrows, highlighted text, pop-ups, etc. Other ways to decrease the differences between experts and novices include reducing unnecessary or redundant information that can distract novices or including text to describe visualizations to help novices process essential content (Gegenfurtner et al., 2011). Similar steps can be taken to help with the age differences by drawing user attention to the most relevant information first.

References

Barfield, W., 1986: Expert-novice differences for software: implications for problem-solving and knowledge acquisition. Behaviour & Information Technology, 5, 15-29.

Chen, S.Y., Fan, J.-P., Macredie, R.D., 2006: Navigation in hypermedia learning systems: experts vs. novices. Computers in Human Behavior, 22, 251-266.

Gegenfurtner, A., Lehtinen, E., Saljo, R., 2011: Expertise Differences in the Comprehension of Visualizations: a Meta-Analysis of Eye-Tracking Research in Professional Domains. Educational Psychology Review, 23, 523-552.

Grahame, M., Laberge, J., Scialfa, C.T., 2004: Age Differences in Search of Web Pages: The Effects of Link Size, Link Number, and Clutter. Human Factors, 46, 385-398.

Hmelo-Silver, C.E., Marathe, S., Liu, L., 2007: Fish Swim, Rocks Sit, and Lungs Breathe: Expert-Novice Understanding of Complex Systems. The Journal of the Learning Sciences, 16, 307-331.

Jacob, R.J.K., Karn, K.S., 2003: Eye Tracking in Human–Computer Interaction and Usability Research: Ready to Deliver the Promises. The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research. Hyona, Radach & Deubel (eds.) Oxford, England.

Jarodzka, H., Scheiter, K., Gerjets, P., van Gog, T., 2010: In the eyes of the beholder: How experts and novices interpret dynamic stimuli. Learning and Instruction, 20, 146-154.

Kastens, K.A., Shipley, T.F., Boone, A.P., Straccia, F., 2016: What Geoscience Experts And Novices Look At, And What They See, When Viewing Data Visualizations. Journal of Astronomy & Earth Sciences Education, 3, 27-58.

Nisbet M. C. (2009). Communicating climate change: Why frames matter to public engagement. Environment, 51, 514-518.

Nisbet, M. C. & Scheufele, D. A. (2009). What's next for science communication? Promising directions and lingering distractions. American Journal of Botany, 96, 1767–1778.

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, 1773-1788.

Perry, S.C., Blanpied, M.L., Burkett, E.R., Campbell, N.M., Carlson, A., Cox, D.A., Driedger, C.L., Eisenman, D.P., Fox Glassman, K.T., Hoffman, S., Hoffman, S.M., Jaiswal, K.S., Jones, L.M., Luco, N., Marx, S.M., McGowan, S.M., Mileti, D.S., Moschetti, M.P., Ozman, D., Pastor, E., Petersen, M.D., Porter, K.A., Ramsey, D.W., Ritchie, L.A., Fitzpatrick, J.K., Rukstales, K.S., Sellnow, T.S., Vaughon, W.L., Wald, D.J., Wald, L.A., Wein, A., and Zarcadoolas, C., 2016, Get your science used—Six guidelines to improve your products: U.S. Geological Survey Circular 1419, 37 p., http://dx.doi.org/10.3133/cir1419.

Pyke, C.R., Bierwagen, B.G., Furlow, J., Gamble, J., Johnson, T., Julius, S., West, J., 2007: A decision inventory approach for improving decision support for climate change impact assessment and adaptation. Environmental Science & Policy, 10, 610-621.

Romano Bergstrom, J.C., Olmsted-Hawala, E.L., Jans, M.E., 2013: Age-Related Differences in Eye Tracking and Usability Performance: Website Usability for Older Adults. International Journal of Human-Computer Interaction, 29, 541-548.

Romsdahl, R.J., 2011: Decision support for climate change adaptation planning in the US: why it needs a coordinated internet-based practitioners' network. Climatic Change, 106, 507-536.

Scannell, L. & Gifford, R. (2013). Personally relevant climate change: The role of place attachment and local versus global message framing in engagement. Environment and Behavior, 45, 60-85.

Simmons, P.E., Lunetta, V.N., 1993: Problem-Solving Behaviors during a Genetics Computer Simulation: Beyond the Expert/Novice Dichotomy. Journal of Research in Science Teaching, 30, 153-173.

Toker, D., Conati, C., Carenini, G., Haraty, M., 2012: Towards Adaptive Information Visualization: On the Influence of User Characteristics. In: Masthoff J., Mobasher B., Desmarais M.C., Nkambou R. (eds) User Modeling, Adaptation, and Personalization. UMAP 2012. Lecture Notes in Computer Science, vol 7379. Springer, Berlin, Heidelberg.

Toker, D., Conati, C., Steichen, B., Carenini, G., 2013: Individual user characteristics and information visualization: connecting the dots through eye tracking. Conference on Human Factors in Computing Systems - Proceedings. 295-304. DOI:10.1145/2470654.2470696.

Wagner, N., Hassanein, K., Head, M., 2014: The impact of age on website usability. Computers in Human Behavior, 37, 270-282.

Table 1. Participant total number, average age, minimum age, maximum age, average total score, minimum total score, maximum total score, and number with previous experience using climate data, divided by degree type (PhD, Master's, Bachelor's, and undergraduate students).

Degree	Number	Avg. Age	Min. Age	Max. Age	Avg. Total Score	Min. Total Score	Max. Total Score	Number with Experience
PhD	6	48	28	65	4.00	3.00	5.00	6
Master's	13	35	24	61	4.77	3.00	6.00	10
Bachelor's	6	31	21	46	4.50	3.00	5.00	2
UG Students	12	20	18	21	4.33	1.00	6.00	9

Table 2. The average TFF and TVD values for novices (undergraduate students) and experts (forestry professionals) for all AOIs related to the tasks and multiple choice questions within the three tools (Extreme MT, SP, and SM/CTMN). The effect size between groups is also included for the TFF and TVD values. Shaded boxes highlight effect sizes with absolute values greater than or equal to 0.5.

AOI Name	TFF Novice Avg.	TFF Expert Avg.	TFF Effect Size	TVD Novice Avg.	TVD Expert Avg.	TVD Effect Size
CNC/Raleigh	14.87	17.46	-0.25	5.30	6.49	-0.24
MT_ColorBarBottom	25.03	75.41	-1.49	4.86	1.79	1.62
MT_Lat/Lon	14.55	36.16	-0.69	4.67	1.10	2.46
MT_LayerOptions	8.74	42.78	-1.03	3.35	0.64	3.65
MT_LeftPanel	85.96	105.86	-0.56	1.81	1.11	0.59
MT_RightPanel	81.41	104.12	-0.44	2.82	2.09	0.24
MT_Timeseries	37.59	80.45	-0.90	14.98	16.98	-0.10
MT_ProjChange	5.57	38.25	-1.01	1.26	0.96	0.39
Columbia	22.12	42.69	-0.64	2.11	3.09	-0.17
SP_Lat/Lon	9.24	34.69	-0.67	4.26	0.96	3.32
SP_LeftPanel	58.20	66.60	-0.28	0.91	0.97	-0.05
SP_ProjAvg	13.93	9.53	0.44	0.98	0.97	0.02
SP_RightPanel	39.69	55.05	-0.58	0.86	2.62	-0.48
SP_Timeseries4	48.90	81.02	-0.66	4.14	4.48	-0.07
SP_Timeseries5	55.66	81.58	-0.56	5.34	5.65	-0.07
Little Rock	44.86	66.02	-0.40	2.63	3.79	-0.12
SM_LayerOptions	19.85	34.35	-0.46	6.85	7.13	-0.04
SM_LeftPanel	38.88	63.46	-0.82	4.71	5.13	-0.06
SM_PlantingPeriod	27.62	81.00	-0.84	3.58	2.30	0.82
SM_RightPanel	43.49	70.06	-0.53	7.39	6.86	0.07

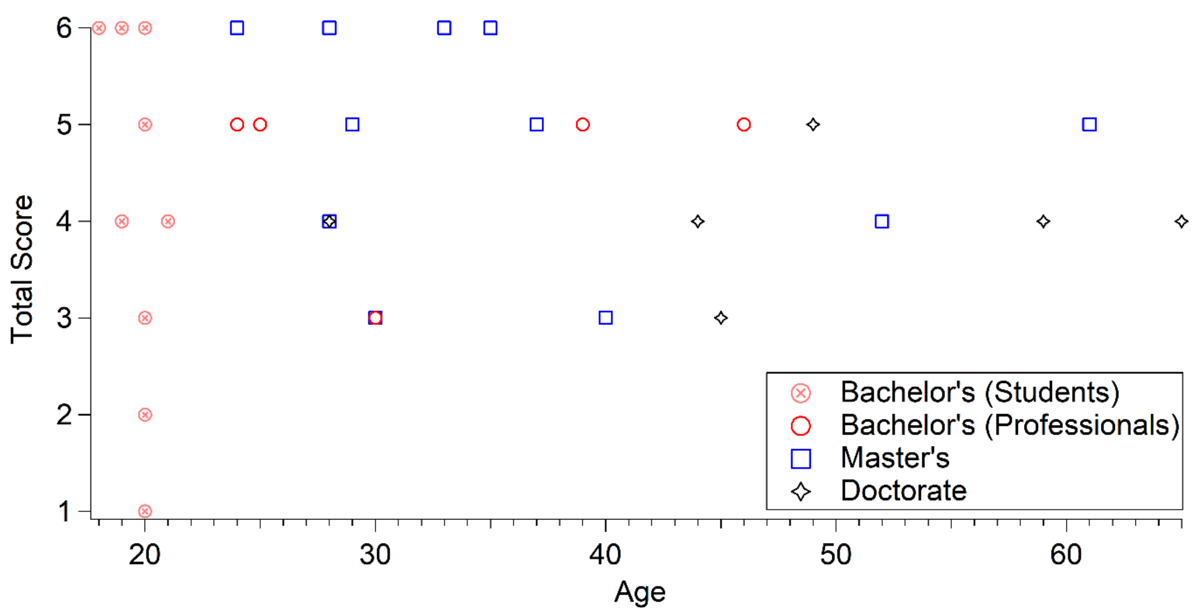


Figure 1. Participants' ages plotted against their total number of correctly answered multiple choice questions, a total score of six being the highest possible. Each marker represents an individual participant, and markers are color-coded by highest degree completed or in progress.

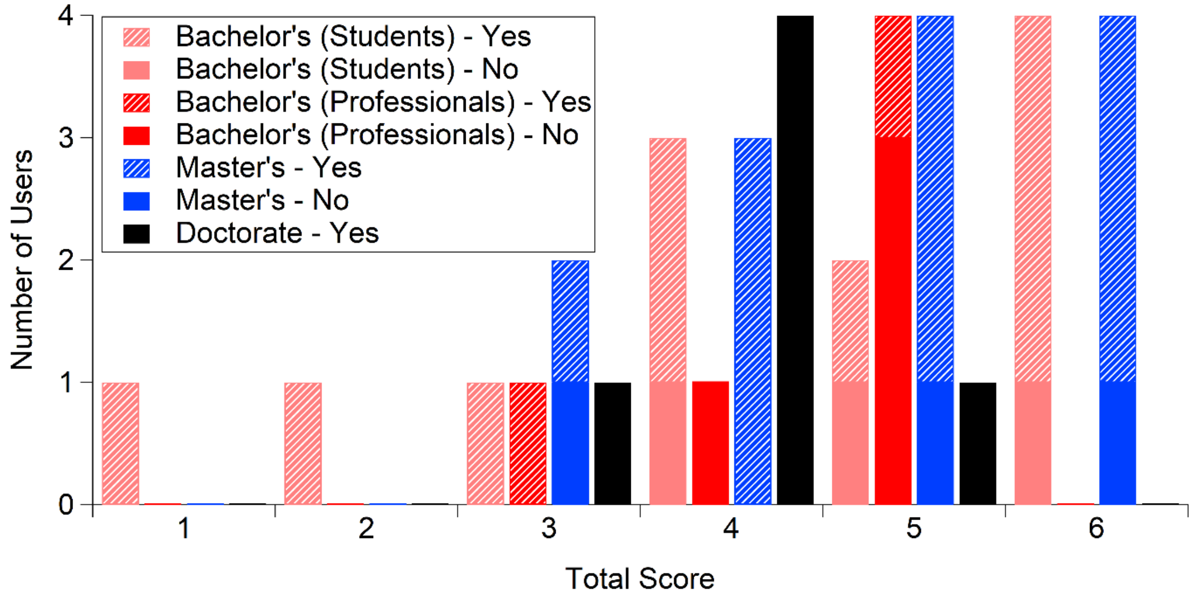


Figure 2. Bar graph of the total score (out of six possible) on the multiple choice questions associated with the three tasks. Users are divided and color-coded by their highest degree achieved or in progress. Those who said they had experience using climate data are denoted by a shaded fill while those who said they did not have experience are denoted by a solid fill.

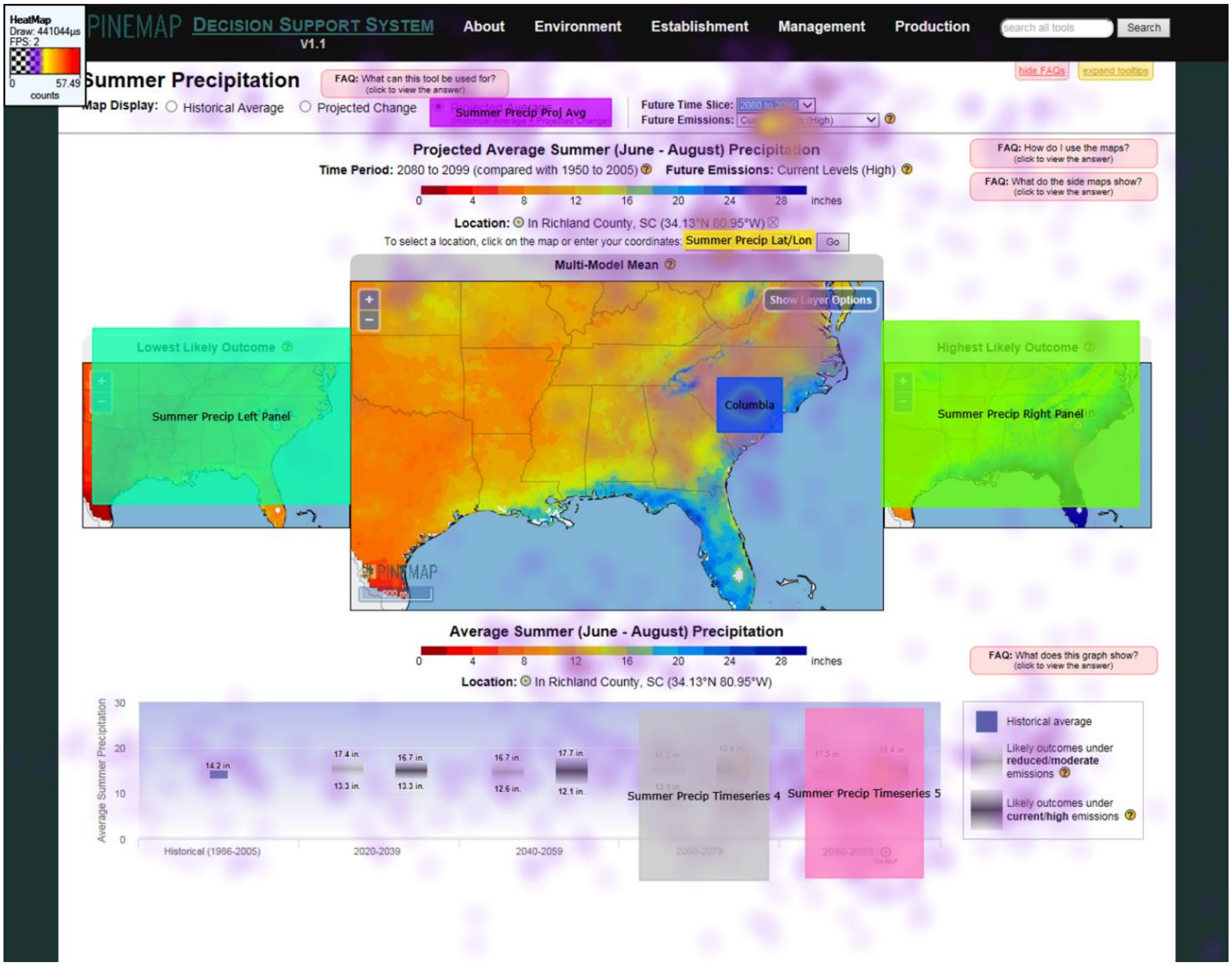
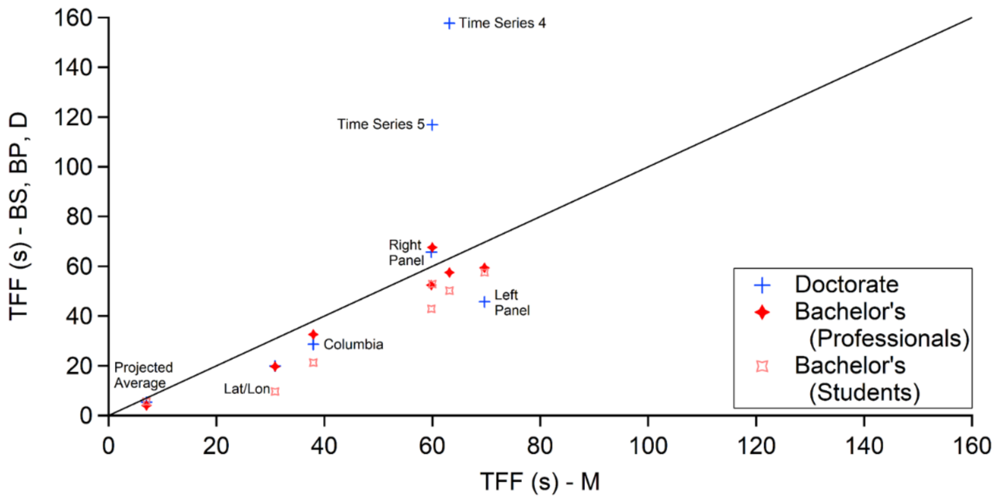


Figure 3. AOIs as shown in the Phase 1 version of the PINEMAP DSS for the SP Tool. AOIs are denoted by colored, slightly transparent boxes that overlay the PINEMAP DSS webpage. Their labels coordinate with the annotations used in the next figure.



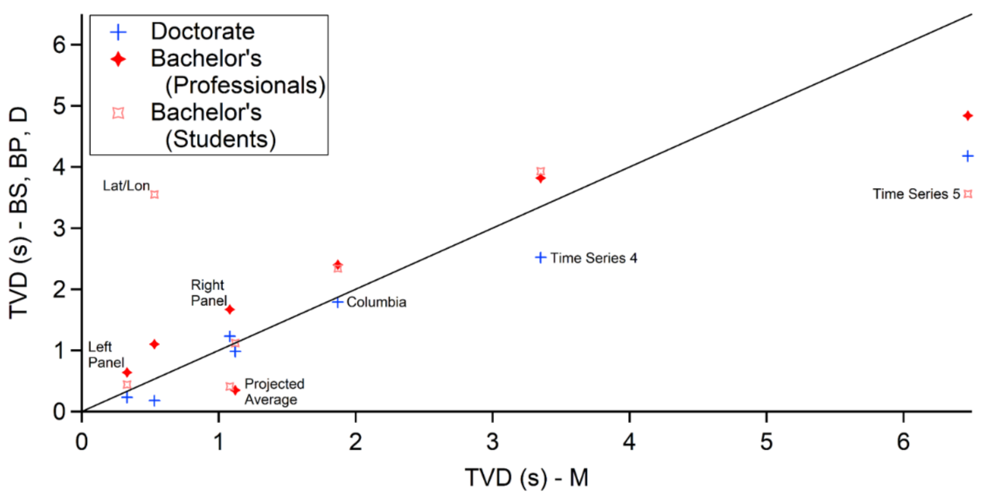


Figure 4. Time to First Fixation (TFF) in seconds (top), and Total Visit Duration (TVD) in seconds (bottom) for the SP Tool. The annotations used in the TFF and TVD plots coordinate with the AOI labels in the previous figure. The TFF and TVD plots both show median values for those with Master's degrees (M) plotted against the median values of those with Doctorate degrees (D), forestry professionals with Bachelor's degrees (BP), and students working toward Bachelor's degrees (BS). The diagonal black line in the TFF and TVD plots is the 1:1 line for comparison. Markers that fall below the line represent cases in which the median values for those with Master's degrees are larger, and markers that fall above the line represent cases in which the median values for those with Master's degrees are smaller.

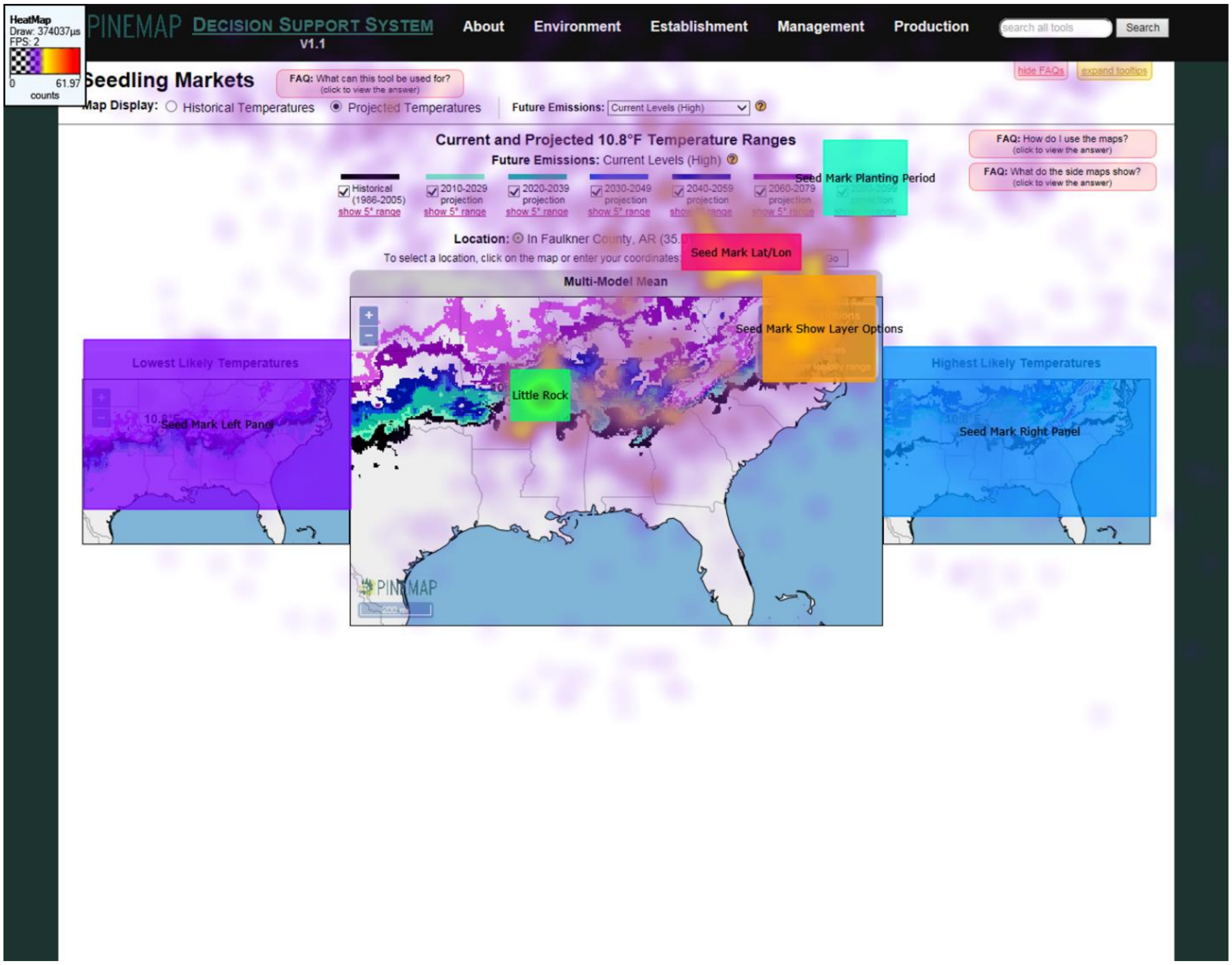


Figure 5. Same as Figure 3 using the SM/CTMN Tool (Phase 1/Phase 2).

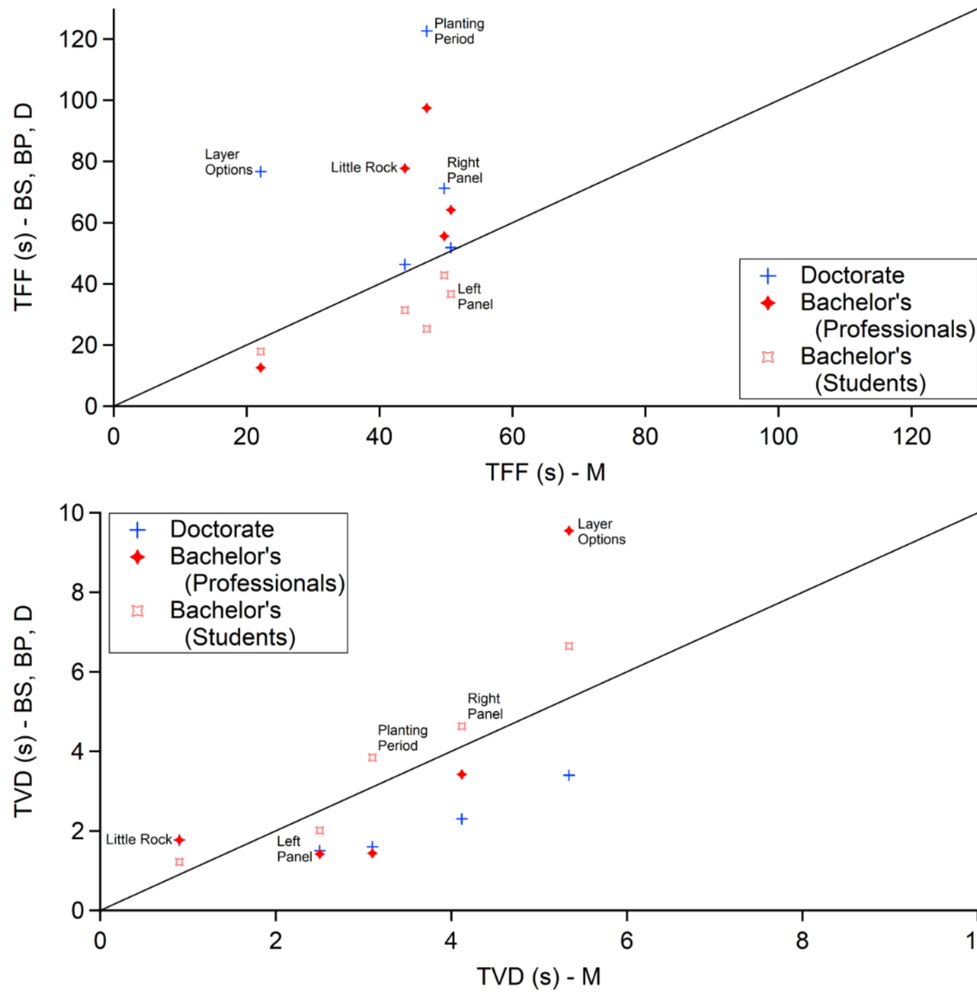


Figure 6. Same as Figure 4 using the SM/CTMN Tool (Phase 1/Phase 2).

SUPPLEMENTAL: Tasks and Multiple Choice Questions

Task: Explore Historical Information

Navigate to the Extreme Minimum Temperature tool using the menu at the top of the page.

Once you're there, select the following options:

Temperature Threshold: 32°F

Map Display: Historical Average

Then answer the following question.

What is a typical value in central North Carolina?

Between 65°F and 75°F

Between 65°C and 75°C

Between 65 days and 75 days

Between 65 years and 75 years

I am not sure.

Task: Explore Projected Change Information

Now select these options:

Map Display: Projected Change

Future Time Slice: 2020-2039

Future Emissions: Current Levels (High)

Location: Raleigh, NC (35.78°N, 78.64°W)

Then answer the following question.

What is the spread in the projected change of extreme minimum temperature occurrence?

about 5 fewer days to about 15 fewer days

about 5°F lower to about 15°F lower

about 5 fewer days to about 25 fewer days

about 5°F lower to about 25°F lower

about 15 fewer days to about 25 fewer days

about 15°F lower to about 25°F lower

I am not sure.

Task: Explore Projected Average Information

Navigate to the Summer Precipitation tool and select the following options:

Map Display: Projected Average

Future Time Slice: 2080-2099

Future Emissions: Reduced Levels (Moderate)

Location: Columbia, SC (34.03°N, 80.90°W)

Then answer the following question.

What is the multi-model mean at this location?

about 13°F

about 13 inches

about 15°F

about 15 inches

about 18°F

about 18 inches

I am not sure.

Task: Explore Time Series Plots

Now examine the time series plot at the bottom of the page.

Then answer the following question.

In the time series plot, which time slice has the greatest spread of possibilities?

Historical

2020-2039 for Reduced (Moderate) Emissions

2040-2059 for Current (High) Emissions

2060-2079 for Reduced (Moderate) Emissions

2080-2099 for Current (High) Emissions

I am not sure.

Task: Explore Projected Temperatures

Navigate to the Seedling Markets tool and select the following options:

Location: Little Rock, AR (34.74°N, 92.30°W)

Map Display: Projected Temperatures

Future Emissions: Current Levels (High)

Then answer the following question.

What is the trend with time from the Historical (black) to the 2080-2099 (magenta) time periods? How do these trends differ from the Lowest Likely to the Highest Likely Temperatures?

Northward with more spread

Southward with more spread

Northward with less spread

Southward with less spread

I am not sure.

Task: Explore a Temperature Range

Next, show the 5° range for the 2080-2099 time period. Also, in the "Show Layer Options" menu, turn on the Native loblolly range layer.

Then answer the following question.

What is the mean projected location of the 5° range relative to the native loblolly range?

It is mostly contained within the native loblolly range.

It is mostly north of the native loblolly range.

It is mostly south of the native loblolly range.

I am not sure.

SUPPLEMENTAL: Questionnaires

Phase 1

1. What sort of historical climate information do you need for your job or research?

2. What sort of future climate information do you need for your job or research?

3. Which websites do you use to find climate information for your job or research?

4. Which climate data or information in the DSS would be most useful for your job or research?

5. Does the PINEMAP DSS provide you with information that you would not be able to find elsewhere? If yes, how so?

6. What additional information would you like included in the PINEMAP DSS to assist you with climate-based decision making for your job or research?

7. What is your age?

8. If you are currently a college student, what is your year, major and current degree in progress (e.g., bachelor's, master's, doctorate)?

9. If you are not currently a college student, what is your highest level of education completed?

10. If applicable, please list your major and degree.

11. If applicable, what is your job title?

12. If applicable, who is your employer?

Phase 2

1. What sort of climate information do you use for your coursework, research, or personal interest?

2. Which websites do you use to find climate information?

3. Which climate data or information in the DSS do you find most interesting?

4. Which of the following tools or activities have you used or been involved with?

    a. Boy Scouts or Girl Scouts

    b. Geocaching

    c. Pokémon Go

    d. Playing video games

    e. Playing computer games

    f. Playing board games

    g. Legos or similar toys

    h. Using Google Maps

    i. Using paper road maps or atlases

    j. Using topographic maps

    k. Using mobile GPS or traffic apps

    l. Treasure hunts or scavenger hunts

5. Over the past year, how much time have you spent doing all of these activities, on average?

a. Less than an hour per day

b. 1-2 hours per day

c. 2-4 hours per day

d. More than 4 hours per day

6. What types of maps do you use most frequently?

7. What is your age?

8. What is your current year in college?

a. Freshman

b. Sophomore

c. Junior

d. Senior

9. What is your college major?

CHAPTER 4: Usability of a Web-based Climate Decision Support System

Lindsay C. Maudlin, Karen S. McNeal, Heather Dinon-Aldridge, Corey Davis, Ryan Boyles, Rachel M. Atkins

[TO BE SUBMITTED FOR PUBLICATION TO JOURNAL OF USABILITY STUDIES]

Abstract

A web-based climate decision support system (DSS) was evaluated using eye tracking to determine which design elements were most salient. After an initial evaluation, the website was changed, and a second phase of evaluation took place to test whether the changes led to an overall increase in usability as measured by efficiency and effectiveness. Both phases of the evaluation included five minutes for users to freely explore the DSS followed by a series of tasks and multiple choice questions to be completed at the users' own paces. The second phase saw an overall increase in efficiency in the average time spent completing the tasks and questions; however, on average, the participants in the first phase were able to answer more questions correctly, suggesting the overall effectiveness was not increased. There was one exception to this: for the last set of tasks and questions, the users in the second phase were able to answer more questions correctly, on average, than those in the first phase. A closer look into this result revealed this increase is dependent on the overall performance of individual users, suggesting usability increases are not universal. Additionally, despite many users finding a specific menu in the free exploration period, they took longer to find and use it during the tasks and questions portion of the study compared to those who did not find it at all during the free exploration

period. These results suggest the need to further study the impacts of cognitive load in ecologically validated studies.

Introduction

Decision support systems have become more popular in the process of disseminating climate information to decision-makers. A decision support system (DSS) is a collection of information used to make a decision located in a common space. DSSs allow all existing climate information to be pared down for a specific group of decision-makers, a type of tailoring of a message (Severin and Tankard, 2001), which is consistent with the recommendations made by Snover et al. (2013) to determine the appropriate sources and types of climate information for the targeted end users.

Once the climate information has been decided upon, two key things to consider before designing and implementing a DSS are the end users and the website design. General website design literature suggests that website designers should recognize that various users come from different backgrounds which can influence the way they will use a website (Morville and Sullenger, 2010). This rings particularly true in the realm of climate change communication where users bring many factors with them (e.g., experience with climate events, political affiliation or ideology) that can influence their willingness to accept the science and take action (Borick and Rabe, 2008; McCright and Dunlap, 2011). For this reason, different users will interact with the same information in different ways, so knowing the audience is an important first step. In addition to knowing the end users and what they bring with them, evaluation is an important next step (Perry et al., 2016) in the climate information dissemination process because it helps to ensure that decision-makers have the information they need to make adaptation decisions in the face of climate change. Unfortunately, DSSs are oftentimes developed but then not evaluated for their overall usability and to ensure they meet the needs of the decision-makers they are serving (Pyke et al., 2007).

To help tackle the website design concerns, the usability field can offer many examples of how usability can be measured and provide various theories to consider and apply. One such theory is the feature-integration theory of attention (Treisman and Gelade, 1980). Within this theory, visual search is an important component which suggests two different phases of attention are utilized when individuals look at and process information (Creager and Gillan, 2016). The first phase involves searching for features such as a particular color or object orientation. Ideally, the target being searched for will stand out in this phase. If not, the target must be detected in the second phase, focal attention, which requires a serial scan of the information. Targets in this category can be obscured by distractors in the field, and the time taken to find a particular target is dependent on the target's ability to stand out from other targets in the first phase. One factor that can influence a target's ability to stand out is how similar the target is to non-targets in terms of their various features (Duncan and Humphreys, 1989). When visual search literature is applied to the evaluation of a DSS, it brings to light the need for critical aspects of the DSS to be detectable in the first phase of visual search. Through the evaluation process, critical aspects can be studied to determine which phase they are detected in, if any. These results can be used to make recommendations for DSS redesign (Rogers and Preston, 2009).

The present study utilized eye tracking to evaluate the usability – defined by efficiency, effectiveness, and satisfaction (Creager and Gillan, 2016) – of a web-based climate DSS, applied usability theory to a real-world application ecological validity (Neisser, 1976; Stork & Schubo, 2010), and aimed to address the following research questions:

1. Which design elements are salient and which are not?

2. Do changes to the web-based climate DSS increase the overall usability? If so, how is the usability increased?

Methods

Study design

A climate DSS was evaluated in two phases (Phase 1 and Phase 2). In both phases, participants were given five minutes to freely explore the DSS. During the free exploration period, participants were able to read about and/or use the DSS tools at their own pace. After the five minutes were up, participants were then given the same series of tasks. There were three major tasks, each of which coordinated with a different DSS tool, but they also included minor tasks within the major task. The major tasks increased in difficulty from the first to the third and served as a way to measure the usability of various DSS features. For each major task and tool combination, there were two multiple choice questions. They were designed to only be answered correctly if the tasks were completed correctly, and they served as a way to determine the extent to which participants understood the material they were interacting with. Throughout this process, participants' eye movements were tracked.

After the eye tracking portion of the study, participants were asked to complete questionnaires. The questionnaires included basic demographic questions (age, gender, level of education) as well as open-ended questions about participants' experiences using climate data in their jobs, research, or courses.

Study populations

Phase 1 participants (n=30) were recruited at a forestry conference in Durham, North Carolina. There were 18 males and 12 females, their average age was 37, and their highest degrees earned or in progress consisted of 6 Bachelor's, 17 Master's, and 6 doctorate degrees (one participant did not respond to this question) in forestry or similar fields. These participants were compensated with their choice of a pre-packaged candy bar or a rain gauge. Phase 2 participants (n=12) were undergraduate students recruited from introductory physical science classes at a large university in the Southeast United States. There were 4 males and 8 females, their average age was 20, and their grade levels were 2 freshmen, 8 sophomores, and 2 juniors in STEM and non-STEM majors. They were compensated with a pre-packaged candy bar.

Climate DSS

The climate DSS evaluated in this study was the Pine Integrated Network: Education, Mitigation, and Adaptation Program (PINEMAP) DSS. The PINEMAP DSS was web-based and designed specifically for foresters in the Southeast United States. It provided important climate information (historical and projected) for the loblolly pine forest industry through interactive map displays.

Three of the PINEMAP DSS tools were used in this evaluation: the Extreme Minimum Temperature (EMT) tool; the Summer Precipitation (SP) tool; and the Seedling Markets (SM) tool. These three were chosen because they represented the two main categories of the PINEMAP tools, Environment (EMT and SP) and Management (SM). The Environment tools focused on the environmental conditions that loblolly pines have historically been exposed to or are expected to be exposed to in the future. The Management tools explored where loblolly pines should be planted based on the climate conditions they will grow in. The Environment tools are more simplistic in the amount and type of information they provide as compared to the more complex Management tools.

Between the two evaluation phases, the PINEMAP DSS was changed slightly. The changes made to the website were based on preliminary findings from Phase 1 and included altering the order of text on the homepage to increase the likelihood that all users read the most important information before leaving the homepage and leaving a map menu open by default so that users were more likely to discover it. Additionally, the SM tool was renamed the Cold-Tolerant Markets for Nurseries (CTMN) tool (Figure 1). The desire to evaluate the helpfulness of these and other changes led to the Phase 2 evaluation.

Tasks and Questions

The three major tasks and related questions (Table 1) were designed for users to follow step-by-step instructions. After following the steps and completing the task, users could then look at the map display(s) and answer the questions. The tasks and questions were given in the same order for all Phase 1 participants. The changes made to the PINEMAP DSS between the phases altered the way in which the SM task and questions could be ordered, so the participants in Phase 2 were given the tasks and questions in the same order as Phase 1 participants for the first four questions, and then the fifth and sixth questions were reversed in order.

Eye tracking

The two phases utilized different eye trackers due to the difference in locations for the data collection portion of the study. Phase 1 used two Tobii X2-60 (60 gaze points per second) eye trackers with portable laptop set-ups, and Phase 2 used one TX-300 (300 gaze points per second) eye tracker with a stationary desktop set-up. To maintain uniform data collection across all participants, each participant went through a calibration process and sat equidistance from the eye tracker (~65 cm). Eye tracking data collected from participants with low calibration rates (<60%) are not used in this study.

After data collection, areas of interest (AOIs) were drawn using the Tobii Studio software. The AOIs were drawn around various aspects of the PINEMAP DSS that were important for completing tasks and answering the related questions. From these AOIs, two eye-tracking metrics were determined and used in this study: time to first fixation (TFF) and total visit duration (TVD). TFF values indicate how eye-catching a particular element is, and TVD values provide insight into which elements a user finds interesting or difficult to interpret.

Analyses

The start and stop times for each user and each task were recorded in addition to the time it took for users to find specific aspects of the PINEMAP DSS (if they did at all). These times were averaged and compared between Phase 1 and Phase 2 using effect size. Eye-tracking metrics were also averaged and compared between Phase 1 and Phase 2 using effect size.

Limitations

Phase 1 and Phase 2 populations are inherently different from each other based on their average ages and education levels. Increasing age has been shown to negatively impact website usability (Romano Bergstrom et al., 2013), and users' education levels combined with work experience in the forestry industry could also increase the likelihood that users have previous experience using and interpreting climate information. Additionally, of the three usability characteristics – efficiency, effectiveness, and satisfaction (Creager and Gillan, 2016) – satisfaction was not measured.

Validity

The study design was developed by geoscience education researchers, and the tasks and multiple choice questions were written by the geoscience education researchers and the PINEMAP DSS developers. Additionally, the entire study was pilot tested by multiple graduate students to identify any confusing or unclear steps.

Results and Discussion

Phase 1 and Phase 2 differences

The average total time spent interacting with the PINEMAP DSS was different between the two phases, as were the average times spent completing the tasks and answering the questions (Table 2). The differences between the two phases led to negative, medium to large effect sizes, where negative values demonstrated that the Phase 2 participants were able to work through the tasks and answer the questions faster than those in Phase 1. Thus, the overall efficiency of the PINEMAP DSS was increased after the website changes were made.

The average total score (out of six possible) was higher in Phase 1 than in Phase 2 (Table 3), suggesting that overall, the effectiveness of the PINEMAP DSS was not increased. However, when the individual questions are compared between phases, those in Phase 1 perform better than those in Phase 2 on the EMT and SP questions, but those in Phase 2 perform better on the SM questions, on average. Thus, with more complex tasks, the Phase 2 group performed better than those in Phase 1, indicating the overall effectiveness of the PINEMAP DSS SM tool was improved after changes were made to the website. There are a few exceptions to this, as noted by examining smaller groups of users based on their overall performance.

High and Low Performers

When considering the gaze plots of various users, similar patterns were found between different groups of users. These groups – High Performers and Low Performers – were determined based on individuals' total scores on all of the multiple choice questions. High Performers are those who scored six out of six, eliminating users who might have had lucky guesses, and Low Performers are those who scored three or less, consistently showing a lesser understanding of the PINEMAP DSS information. In Phase 1, there were four users (out of 30) who earned a perfect score. In Phase 2, there were also four users (out of 12) who earned a perfect score.

When gaze plots are compared between High Performers in Phase 1 and Phase 2 (Figure 2 and Figure 3, respectively) for the SM tool, there are many similarities. Both groups fixated on color bars and various features that can be turned off and on at the top of the page. They also looked at all three of the map panels, something that was important for answering both questions correctly. However, there was one key difference between the groups of High Performers – the time it took them to complete the entire SM tool evaluation. The average time for the four High Performers from Phase 1 to complete the tasks and answer both questions was 212 seconds, while the average time from Phase 2 was only 117 seconds. Both groups were equally successful in completing the tasks and answering the questions correctly, but those in Phase 2 were able to do so more quickly than those in Phase 1. This indicated that the PINEMAP DSS design changes made it more efficient for High Performers to complete the tasks and successfully answer the related questions on the more complex SM tool.

There were dramatic differences in the gaze plots between the Phase 2 High Performers and Low Performers (Figure 3 and Figure 4, respectively). Low Performers did not fixate in as many different areas as did the High Performers, particularly the color bars and other elements at the top of the page. Most important for answering the questions, the Low Performers did not fixate at all on the left and right map panels, with the exception of a few fixations from one user. This is consistent with previous work (Tsai et al., 2012). By failing to look at and interpret the left and right map panels, these users failed to interact with the information needed to correctly answer the questions. Another key difference between the Phase 2 High Performers and Low Performers was the total time spent on the SM tool evaluation. Low Performers only spent an average of 87 seconds on this tool (compared to 117 seconds of High Performers). Based on these findings, Low Performers did not find nor recognize the importance of certain information included in the PINEMAP DSS to complete the tasks correctly, nor did they spend the time needed to interpret this additional information.

Finding DSS Features

The “Layer Options” menu was initially minimized by default in Phase 1, and users struggled to find it due to its small size and similar coloring to the map in the background. The percentage of those who found this menu and clicked on an option in it during the free exploration period was 35%, and six users in Phase 1 never found it, even during the tasks and questions portion of the study. For this reason, it was left open by default in Phase 2 (Figure 1).

The percentage of Phase 2 users who found the “Layer Options” menu and clicked on an option in it during the free exploration period was 58%, showing an increase in the discoverability of the menu compared to Phase 1. Despite having a higher rate of users who found this feature during the free exploration period, Phase 2 users took, on average, ~21 seconds longer to find and use the “Layer Options” menu to turn on the “Native loblolly range” during the tasks and questions portion of the study than Phase 1 users (Table 4). The average time taken to find and use this feature was compared to two eye-tracking metrics, TFF and TVD, calculated based on the “Layer Options” menu AOI (Table 4). On average, Phase 2 participants had shorter TFF values than Phase 1 participants (an effect size of -0.46) but relatively similar TVD values (an effect size of -0.04). The shorter TFF values but longer time to find and use the menu suggested that the issue for Phase 2 users was not in the discoverability or findability of the menu but rather in their understanding of the menu’s role in completing the major task and answering the related questions in the SM tool.

Additionally, of all those who found and used the menu during the free exploration period from both Phase 1 and Phase 2, there was an increase in the average number of seconds it took them to locate the same menu during the tasks compared to those who had not found and used it initially in the free exploration (63 seconds compared to 36 seconds with an effect size of 0.93). The effect size was even more dramatic for users in Phase 2 (1.38 based on 79 seconds for those who had previously found and used the menu compared to 51 seconds for those who had not). This was an unexpected result and led to more questions than answers as to why users struggled to find a familiar object. One possible explanation is that users were cognitively overloaded (Mayer and Moreno, 2003) by this point in the study, thus making a relatively easy task harder, but further research is needed.

Conclusion and Future Work

Differences were found between groups of users who had used the PINEMAP DSS before and after changes were made to the website. Phase 1 users, on average, took more time to complete the tasks and answer the related questions, but they were able to answer more questions correctly, on average.

Gaze plots showed similar patterns between High Performers in both Phase 1 and Phase 2; however, those in Phase 2 were able to move through the tasks and questions roughly 100 seconds fasters than those in Phase 1. Conversely, gaze plots revealed differences between Phase 2 High Performers and Low Performers where Low Performers did not look at the relevant maps to answer the questions. One recommendation is that DSS development should be tailored for the end users such that Low Performers can perform like High Performers when they interact with a DSS designed with them in mind. Examples include drawing more attention to the left and right map panels or other DSS elements that are necessary for successful interaction.

These findings suggest that, in general, the PINEMAP DSS usability increased for Phase 2 users in regard to their efficiency, but not necessarily in regard to their overall effectiveness. When examining High and Low Performers, the overall usability became even more evident in regard to efficiency and effectiveness differences, where the overall usability of the PINEMAP DSS was increased between Phase 1 and Phase 2 for High Performers, but Low Performers still struggled despite the design changes.

Future work should include additional studies to validate visual search and usability theories and increase ecological validity (Neisser, 1976). Additionally, future work should address the impacts that cognitive load might potentially have on the findability of elements during a search task that users have previously viewed such as the "Layer Options" menu in this study.

References

Borick, C. P. and Rabe, B. G. (2008). A reason to believe: Examining the factors that determine Americans' views on global warming. Issues in Governance Studies, 18, 1-14.

Creager, J. H., & Gillan, D. J. (2016). Toward understanding the findability and discoverability of shading gradients in almost-flat design. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 60, 339–343.

Duncan, J., Humphreys, G. W., 1989: Visual Search and Stimulus Similarity. Psychological Review, 96, 433-458.

Mayer, R. E., and Moreno, R., 2003: Nine Ways to Reduce Cognitive Load in Multimedia Learning. Educational Psychologist, 38, 43-52.

McCright, A. M., and Dunlap, R. E. (2011). The politicization of climate change and polarization in the American public's views of global warming, 2001-2010." The Sociological Quarterly, 52, 155-194.

Morville, P., Sullenger, P., 2010: Ambient Findability: Libraries, Serials, and the Internet of Things. The Serials Librarian, 58, 33-38.

Neisser, U.: Cognition and Reality: Principles and Implications of Cognitive Psychology, WH Freeman, New York, 1976.

Perry, S.C., Blanpied, M.L., Burkett, E.R., Campbell, N.M., Carlson, A., Cox, D.A., Driedger, C.L., Eisenman, D.P., FoxGlassman, K.T., Hoffman, S., Hoffman, S.M., Jaiswal, K.S., Jones, L.M., Luco, N., Marx, S.M., McGowan, S.M., Mileti, D.S., Moschetti, M.P., Ozman, D., Pastor, E., Petersen, M.D., Porter, K.A., Ramsey, D.W., Ritchie, L.A., Fitzpatrick, J.K., Rukstales, K.S., Sellnow, T.S., Vaughon, W.L., Wald, D.J., Wald, L.A., Wein, A., and

Zarcadoolas, C., 2016, Get your science used—Six guidelines to improve your products: U.S. Geological Survey Circular 1419, 37 p., http://dx.doi.org/10.3133/cir1419.

Pyke, C.R., Bierwagen, B.G., Furlow, J., Gamble, J., Johnson, T., Julius, S., West, J., 2007: A decision inventory approach for improving decision support for climate change impact assessment and adaptation. Environmental Science & Policy, 10, 610-621.

Rogers, R., & Preston, H. (2009). Usability analysis for redesign of a Caribbean academic library web site: a case study. OCLC Systems & Services: International Digital Library Perspectives, 25, 200–211.

Romano Bergstrom, J.C., Olmsted-Hawala, E.L., Jans, M.E., 2013: Age-Related Differences in Eye Tracking and Usability Performance: Website Usability for Older Adults. International Journal of Human-Computer Interaction, 29, 541-548.

Severin, W.J., Tankard, Jr., J.W., 2001: Communication theories: Origins, methods, and uses in the mass media (5th ed.). New York: Addison Wesley Longman. Ch. 1.

Snover, A.K., Mantua, N.J., Littell, J.S., Alexander, M.A., McClure, M.M., Nye, J., 2013: Choosing and using climate-change scenarios for ecological-impact assessments and conservation decisions. Conservation Biology, 27, 1147-1157.

Stork, S., Schubo, A., 2010: Human Cognition in manual assembly: Theories and applications. Advanced Engineering Informatics, 24, 320-328.

Tsai, M.-J., Hou, H.-T., Lai, M.-L., Liu, W.-Y., Yang, F.-Y., 2012: Visual attention for solving multiple-choice science problem: An eye-tracking analysis. Computers & Education, 58, 375-385.

Treisman, A.M., Gelade, G., 1980: A Feature-Integration Theory of Attention. Cognitive Psychology, 12, 97-136.

Table 1. The tasks and multiple choice questions organized by PINEMAP DSS tool. Q1 refers to Question 1, and Q2 refers to Question 2.

EXTREME MINIMUM TEMPERATURE TOOL	Task	Navigate to the Extreme Minimum Temperature tool.
		Select the following options:
		Temperature Threshold: 32°F Map Display: Historical Average
	Q1	What is a typical values in central North Carolina?
		Between 65°F and 75°F Between 65°C and 75°C Between 65 days and 75 days
		Between 65 years and 75 years I am not sure
	Task	Now select these options:
		Map Display: Projected Change Future Time Slice: 2020-2039 Future Emissions: Current Levels (High)
		Location: Raleigh, NC (35.78°N, 78.64°W)
	Q2	What is the spread in the projected change of extreme minimum temperature occurrence?
		about 5 fewer days to about 15 fewer days about 5°F lower to about 15°F lower
		about 5 fewer days to about 25 fewer days about 5°F lower to about 25°F lower
		about 15 fewer days to about 25 fewer days about 15°F lower to about 25°F lower I am not sure
SUMMER PRECIPITATION TOOL	Task	Navigate to the Summer Precipitation tool.
		Select the following options:
		Map Display: Projected Average Future Time Slice: 2080-2099
		Future Emissions: Reduced Levels (Moderate) Location: Columbia, SC (34.03°N, 80.90°W)
	Q1	What is the multi-model mean at this location?
		about 13°F about 13 inches about 15°F about 15 inches
		about 18°F about 18 inches I am not sure
	Task	Now examine the time series plot at the bottom of the page.
	Q2	In the time series plot, which time slice has the greatest spread of possibilities?
		Historical 2020-2039 for Reduced (Moderate) Emissions 2040-2059 for Current (High) Emissions
		2060-2079 for Reduced (Moderate) Emissions 2080-2099 for Current (High) Emissions I am not sure
SEEDLING MARKETS TOOL	Task	Navigate to the Seedling Markets tool.
		Select the following options:
		Location: Little Rock, AR (34.74°N, 92.30°W) Map Display: Projected Temperatures
		Future Emissions: Current Levels (High)
	Q1	What is the trend with time from the Historical (black) to the 2080-2099 (magenta) time periods? How do these trends differ from Lowest Likely to the Highest Likely Temperatures?
		Northward with more spread Southward with more spread
		Northward with less spread Southward with less spread I am not sure
	Task	Show the 5° range for the 2080-2099 time period.
		In the "Show Layer Options" menu, turn on the Native loblolly range layer.
	Q2	What is the mean projected location of the 5° range relative to the native loblolly range?
		It is mostly contained within the native loblolly range It is mostly north of the native loblolly range
		It is mostly south of the native loblolly range I am not sure

Table 2. The average time spent on each question for all participants (Average), participants in Phase 1 (Phase 1), participants in Phase 2 (Phase 2), the standard deviation (Std. Dev.) used to calculate the effect size, and the effect size (Effect Size). The last column includes the same information for the total time (including the five minutes of free exploration). All units are in seconds. The questions are labelled by tool – Extreme Minimum Temperature (EMT), Summer Precipitation (SP), and Seedling Markets (SM) – and by question number – Question 1 (Q1) and Question 2 (Q2). The times for SM Q1 and SM Q2 are combined to eliminate any biases that might exist based on the order participants interacted with those questions between Phase 1 and Phase 2.

	EMT Q1	EMT Q2	SP Q1	SP Q2	SM Q1&Q2	Total Time
Average	100.03	102.21	85.53	52.71	201.79	860.42
Phase 1	105.88	115.23	94.27	61.12	225.00	919.35
Phase 2	87.33	74.00	66.58	34.50	151.50	732.75
Std. Dev.	34.95	38.84	48.72	34.40	78.58	191.46
Effect Size	-0.53	-1.06	-0.57	-0.77	-0.94	-0.97

Table 3. For each question answered correctly, participants were given one point. The total score was the sum of the points for all six questions. The average score on each question for all participants (Average), participants in Phase 1 (Phase 1), participants in Phase 2 (Phase 2), the standard deviation (Std. Dev.) used to calculate the effect size, and the effect size (Effect Size). The last column includes the same information for the total score (six was the highest possible). The questions are labelled by tool – Extreme Minimum Temperature (EMT), Summer Precipitation (SP), and Seedling Markets (SM) – and by question number – Question 1 (Q1) and Question 2 (Q2).

	EMT Q1	EMT Q2	SP Q1	SP Q2	SM Q1	SM Q2	Total Score
Average	0.89	0.47	0.84	0.76	0.68	0.82	4.47
Phase 1	0.92	0.54	0.88	0.77	0.62	0.81	4.54
Phase 2	0.83	0.33	0.75	0.75	0.83	0.83	4.33
Std. Dev.	0.27	0.50	0.32	0.42	0.49	0.39	0.93
Effect Size	-0.34	-0.41	-0.42	-0.05	0.45	0.07	-0.22

Table 4. The time to first fixation (TFF), the total visit duration (TVD), and the total time spent looking for the native loblolly pine range (Loblolly) in seconds. The average time for all participants (Average), participants in Phase 1 (Phase 1), participants in Phase 2 (Phase 2), the standard deviation (Std. Dev.) used to calculate the effect size, and the effect size (Effect Size).

	TFF	TVD	Loblolly
Average	29.77	7.05	47.68
Phase 1	34.35	7.13	38.62
Phase 2	19.85	6.85	67.33
Std. Dev.	31.49	6.55	29.46
Effect Size	-0.46	-0.04	0.97

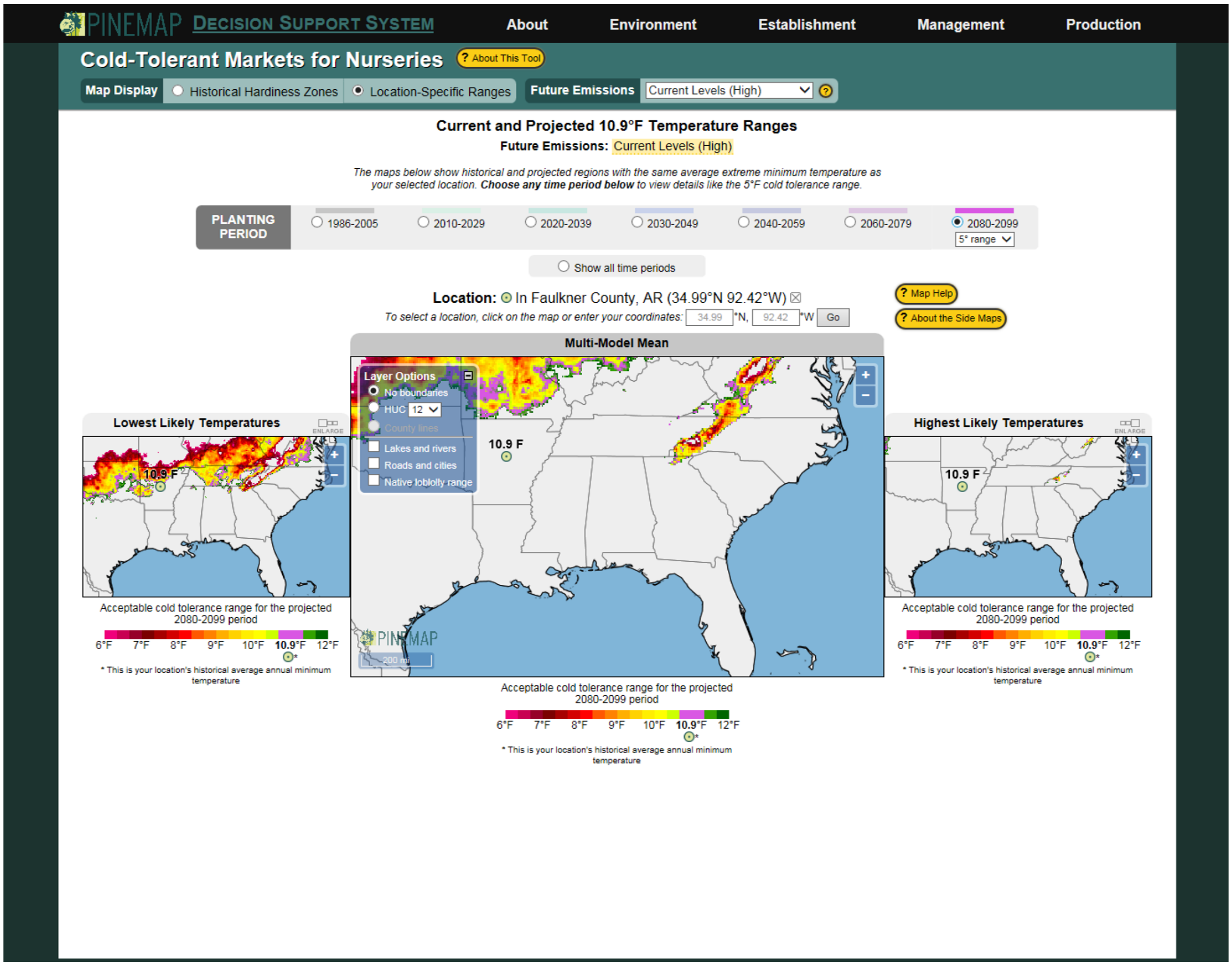


Figure 1. The Cold-Tolerant Markets for Nurseries tool (Phase 2). The “Layer Options” menu is open by default and located in the upper left corner of the center map.

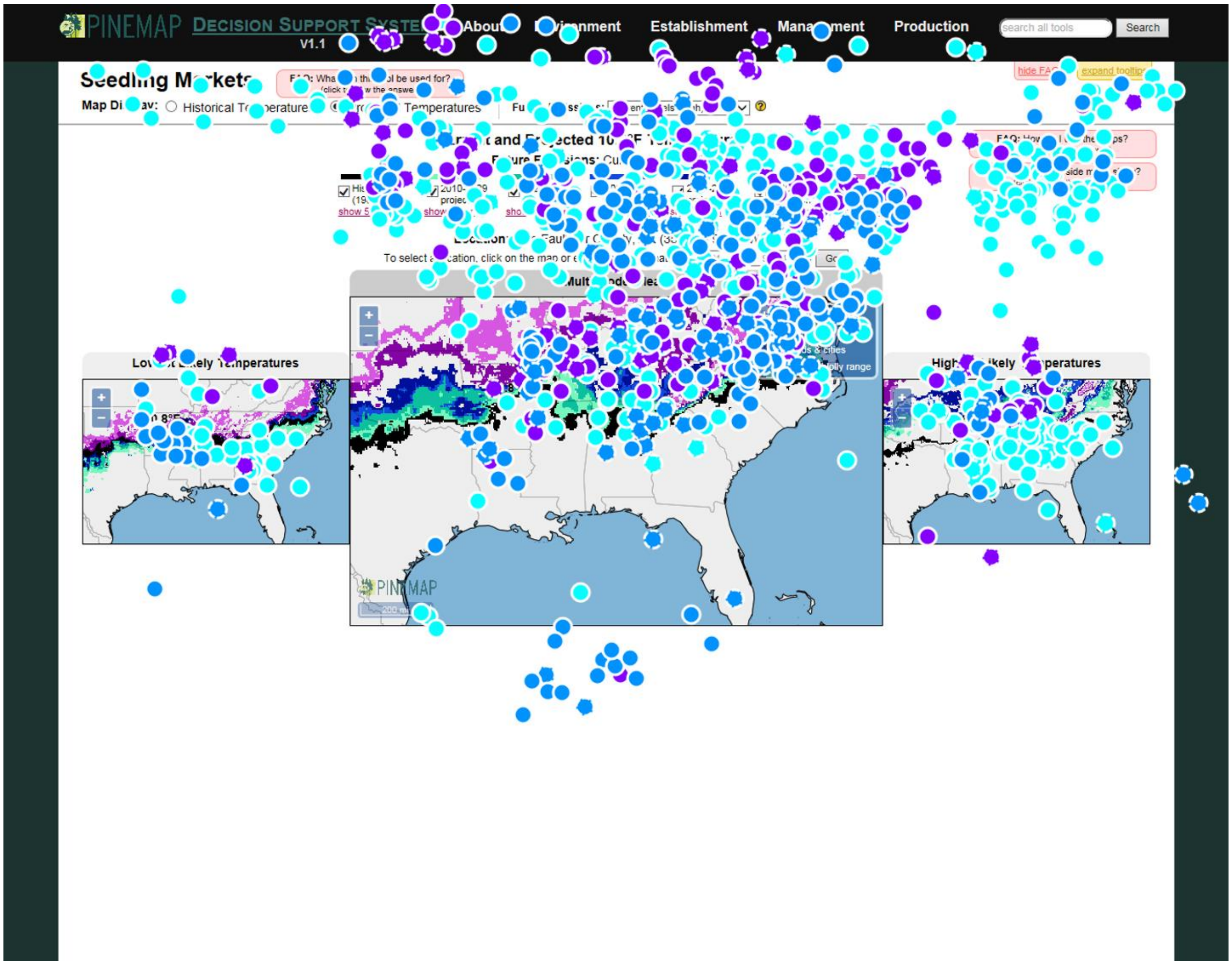


Figure 2. Phase 1 High Performers (n=3). Each color represents a separate user. Markers represent individual fixation locations. A fourth user from Phase 1 also was considered a High Performer but is not included in this figure due to limitations in the Tobii software.

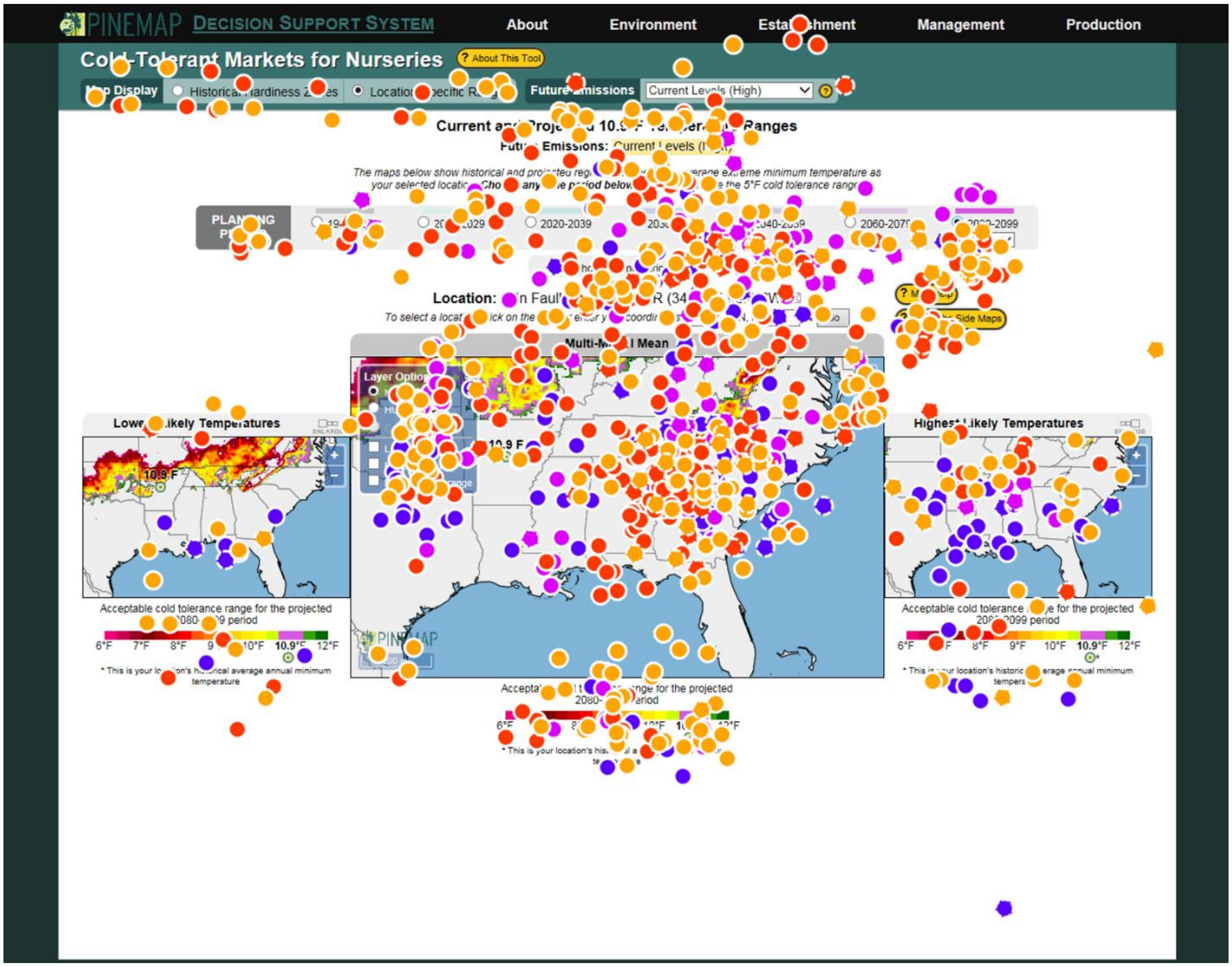


Figure 3. Phase 2 High Performers (n=4). Each color represents a separate user. Markers represent individual fixation locations.

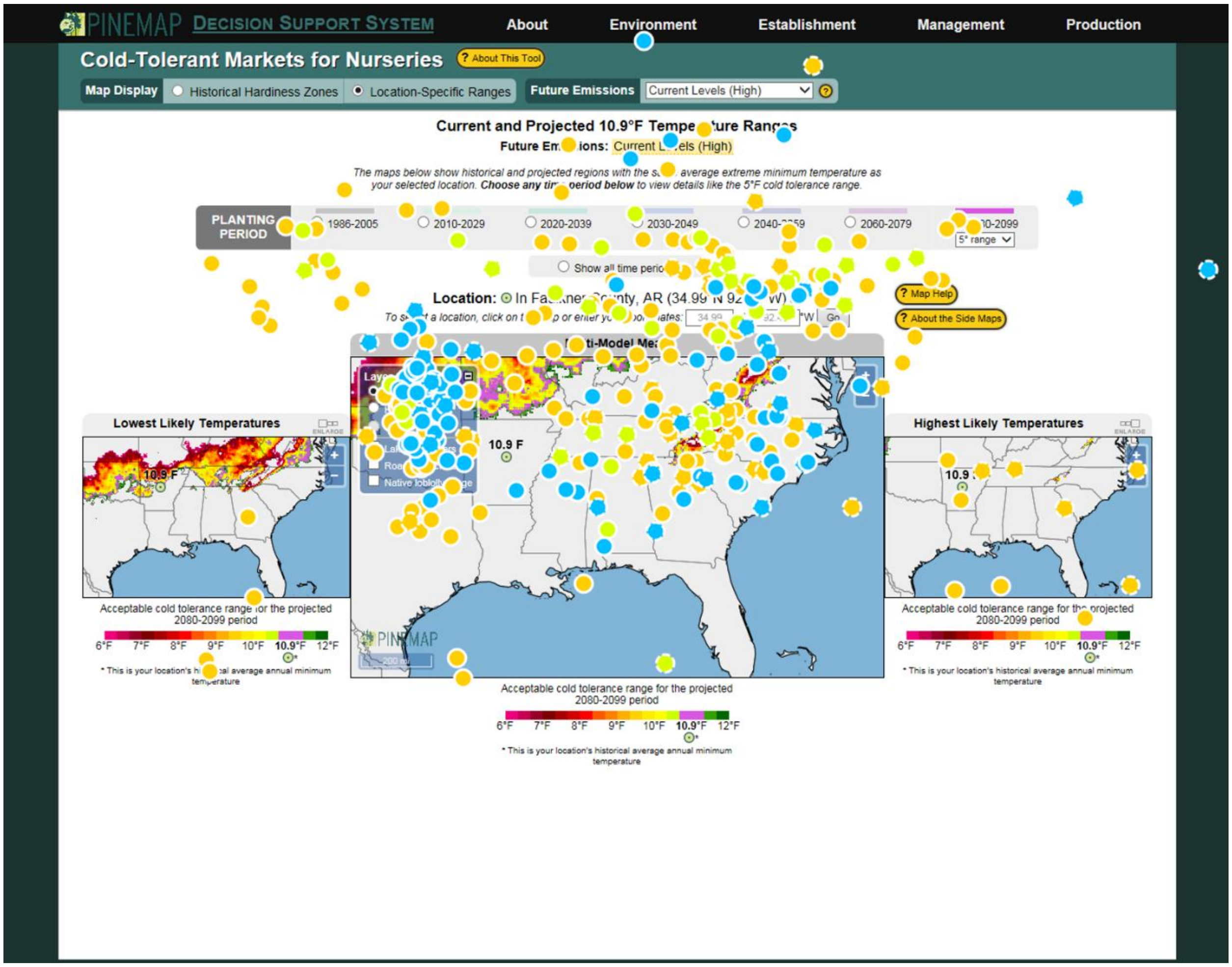


Figure 4. Phase 2 Low Performers (n=3). Each color represents a separate user. Markers represent individual fixation locations.

CHAPTER 5: Conclusion

Summary of Key Findings

- Males and females exhibit different eye-tracking behaviors during both the free exploration period and the tasks and multiple choice questions portions of the study.
- Females tend to have more thorough fixation patterns and tend to read through the text more than males do.
- These behaviors impact performance on related questions. Males correctly answer more questions than females, on average.
- The three lowest performers in the entire study were females. Although spatial ability and exposure to spatial activities do not directly predict overall performance, these three females had lower than average spatial ability scores and less experience with spatial activities.
- Experience, as defined by education level and exposure to climate information, is not an accurate predictor of performance on related tasks and multiple choice questions.
- Age is a confounding factor as a result of the DSS being web-based.
- The overall efficiency increased after changes were made to the DSS, as measured by the decrease in time spent completing the tasks and answering the related multiple choice questions.
- The overall effectiveness did not increase after changes were made to the DSS, as measured by user accuracy in answering the multiple choice questions. The exception is that users in the second evaluation were able to answer more questions correctly in the final set of tasks and questions, thus suggesting that the effectiveness did increase within this portion of the DSS.

- The gaze plots of high and low performers are drastically different, suggesting that while the overall efficiency and some aspects of the effectiveness of the DSS increased after changes were made, not all users will benefit from these changes.
- A particular menu that was closed by default in the first evaluation and opened by default in the second evaluation led to a series of unexpected results. First, more users found the menu during the free exploration period in the second evaluation than in the first evaluation.
- Of those who did find the menu during the free exploration period, they took longer, on average, to find and use the menu during the tasks and questions portion of the study, suggesting that users might have been cognitively overloaded.

Contributions to the Field

This study evaluates a DSS through the use of eye tracking to determine the factors that serve as obstacles in the usability and understandability of the DSS, contributing to the field by addressing gaps in the existing literature and raising new questions for future investigations. The key findings from this study add to the body of literature on gender differences in website usability studies, expert-novice differences within map reading and within a website usability context, and overall usability of a web-based DSS.

This evaluation is the first of its kind and provides insight into the gender differences that exist as males and females interact with a web-based climate DSS and interpret climate information from maps. Eye tracking has highlighted the quantitative differences between males and females in their search and navigation behaviors, and multiple choice questions demonstrate that the differences in eye-tracking behaviors have direct consequences on a person's success in correctly answering questions about the content. This raises the question: If the website were designed such that males and females were guaranteed to interact with the same content in the same order, would there still be differences in overall performance? If no, this suggests that these differences are solely a usability concern.

One possible explanation for these differences relates to a person's spatial ability and exposure to spatial activities that can help improve spatial ability. In this study, there is not a strong, direct link between spatial ability and overall performance; however, the three lowest performers were all females and had lower than average spatial abilities and less exposure to spatial activities. This warrants further investigation into the role that spatial ability, best measured with multiple instruments, and spatial activities have in developing the skillset needed to navigate and correctly interpret web-based climate information.

The expert-novice differences in this study were unanticipated. Education level and familiarity with similar content were hypothesized to be accurate predictors of overall performance; however, age was a confounding factor in this study that masked the role education level and previous experience with climate information had in overall performance. Usability of this and other DSSs would be increased if the end-users were kept in mind during the development stage and included in the evaluation process to ensure that age and other user characteristics are accounted for and addressed in website design. The unanticipated impact that age had in this study is reason to investigate whether other user characteristics have similar masking effects in this and other studies.

The improvement in efficiency between both phases of this study demonstrates that usability can be altered through changes made to a website after an initial evaluation. The increased effectiveness of only one part of the website suggests that more improvements could be beneficial but also leads to the following questions:

1. What was different between the part of the website that saw increased effectiveness and the parts that did not?
2. Although the average effectiveness increased for this part of the website, a few users did not benefit from the website changes between phases. What prevented them from experiencing the same increased effectiveness that their peers did?

Another unanticipated result from this study was the disparity in time taken to find a particular menu between groups of those who had found it initially during the free exploration portion of the study and those who had not. One possible explanation is that those who had found the menu previously became cognitively overloaded between the initial discovery and when they actually needed to use the menu later in the study. This warrants further study of cognitive loads of users as they navigate through and interpret climate information in a web-based format.