

Evaluating The Three Dimensions of Usability for an Online Climate Modeling Tool: A
Coproducted Eye Tracking Study

by

Jena (Ally) Brown

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
August 5th, 2023

Geoscience education, climate change, usability, eye tracking, coproduction, mixed methods

Copyright 2023 by Jena (Ally) Brown

Approved by

Dr. Karen S. McNeal, Chair, Associate Professor of Geosciences

Dr. Chandana Mitra, Associate Professor of Geosciences

Dr. Mark Chandler, Center for Climate Systems Research, Columbia University & NASA/GISS

ABSTRACT

Science tools provide educators with a method to teach their students complex issues like climate change, however, there is still much work to be done to ensure that science tools are efficient and usable for teachers and students, alike. This study aimed to assess the usability of two versions of an online global climate modeling webtool, known as EzGCM, each with intentional design changes that were rooted in the current understanding of user design frameworks. Usability was assessed by comparing efficiency, accuracy, and satisfaction of student participants between versions of the website using a blend of qualitative and quantitative data including pre-post surveys, eye tracking data, task correction, and retrospective interviews. The findings in this study suggest that the redesigned version of EzGCM had a higher usability than the original version when comparing many of the changes made between versions. Beyond EzGCM, this study stresses the importance of usability testing for new pedagogical tools using this three-dimensional approach and makes recommendations for other website designers regarding page layout and website element design based on the findings.

Acknowledgments

I would like to extend my deepest appreciation to my advisor, Dr. Karen McNeal, and my colleagues in the Geocognition Laboratory for supporting the development of this project. In addition, I would like to thank Dr. Mark Chandler for giving me the opportunity to collaborate on his website, EZGCM. Finally, I would like to extend my gratitude to Dr. Mitra who has agreed to sit on my committee and help with the curation of this document.

This material is based upon work supported by the National Science Foundation under Grant Number (NSF Grant #1922687). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Table of Contents

Abstract	2
Acknowledgments.....	3
List of Tables	7
List of Figures	8
List of Abbreviations	9
Introduction.....	10
Visal Attention and Website Design.....	12
Eye Tracking for Usability.....	13
Research Questions	14
Materials and Methods	14
Coproducton Process	15
Development of A & B Designs	15
Participant Task	20
Pilot and Recruitment	21
Intervention with Eye tracking.....	22
Post survey and Interview	24
Analysis.....	25
Data Preparation	26
Areas of Interest.....	26
Excel Preparation	31
Quantitative and Statistical Analysis	32

Efficiency	33
Accuracy	33
Satisfaction.....	37
Qualitative Analysis.....	37
Interviews	37
Heat Maps and Gaze Plots	38
Results.....	38
Quantitative Results	38
Efficiency.....	38
Accuracy	40
Satisfaction.....	41
Qualitative Results	43
Interviews	43
Heat Maps and Gaze Plots	44
Discussion.....	52
Run Simulations Page.....	52
Proceed to Post Processing Button	52
Pause/Play Button	53
Simulations Menu	54
Post Processing Page.....	55
Instructional Detail.....	55
Proceed to Visualizations Button.....	56
Visualizations Page.....	56

Create Map/Plot Button	56
Operations Menu.....	57
Overall Satisfaction	58
Conclusions and Recommendations	59
Limitations and Future Directions	61
Significance.....	62
References	63
Appendix A.....	66
Appendix B.....	70
Appendix C.....	73

List of Tables

Table 1	20
Table 2	25
Table 3	26
Table 4	40
Table 5	41
Table 6	41
Table 7	42
Table 8	43

List of Figures

Figure 1	16
Figure 2	17
Figure 3	18
Figure 4	28
Figure 5	29
Figure 6	30
Figure 7	34
Figure 8	36
Figure 9	46
Figure 10	47
Figure 11	48
Figure 12	50
Figure 13	51

List of Abbreviations

EzGCM	Easy Global Climate Model website
GCM	Global Climate Model
NASA	National Aeronautics and Space Administration
NGSS	Next Generation Science Standards
RS	Run Simulations
PP	Post Processing
Vis	Visualizations
IPCC	International Panel on Climate Change
IRB	Institutional Review Board
AOI	Areas of Interest
Avg	Average
Std dev	Standard Deviation
df	Degrees of Freedom

INTRODUCTION

With the exponential advancement of technology in recent years, educators have become increasingly dependent on educational web tools to teach their students with engaging content (Yuen et al., 2011). In science education in particular, many internet-based tools have been developed to support inquiry-based practices by having students investigate natural phenomena and generate practical solutions to real world issues (Kim et al., 2007). Science tools can be especially important when teaching complex problems like climate change because it involves conceptualizing the complex interactions within the Earth systems on various spatial and temporal scales which is often difficult to conceptualize (McNeal et al., 2014).

Scientists visualize and interpret these complex interactions by using tools like climate models which can be used to simulate the climate consequences of an increase in atmospheric CO₂. Many climate models predict changes in mean global climate both in instantaneous doubling of CO₂ equilibrium simulations and time-dependent simulations in which the CO₂ concentration is increased incrementally over several model years (Shackley et al, 1998). However, the tools and models that scientists use are rarely translated into the classroom for students due to their difficulty and requirement for intense computational power.

The use of models affords students an opportunity to overcome abstractness in climate concepts by engaging with evidence in order to establish understanding derived from their own inquiries (McNeal et al., 2014; Develaki, 2017). Models in the classroom allow students to engage in evidence-based reasoning using existing climate data to make scientific inquiries and conclusions, which is a key characteristic of a scientifically literate person (Develaki, 2017; Bhattacharya et al., 2021). Although the Next Generation Science Standards (NGSS) have emphasized the importance of teaching scientific practices and essential thinking skills, such as

modeling, there are few resources available for educators to do so (McNeal et al., 2014; Bhattacharya et al., 2021). For this reason, science education tools like Easy Global Climate Model (EzGCM), an authentic, data-driven, computer-based climate modeling toolkit, was created to make modeling accessible to students and encourage evidence-based reasoning to improve climate literacy (Brown et al., 2010). EzGCM is a gateway toolkit that can help novice students better understand Global Climate Change and make sense of the major national and international climate assessments that inform and guide climate change policy (Bhattacharya, 2020). Students can make robust data-driven and model-based predictions about future climate change using this web-based teaching tool. When combined with a better understanding of Earth's climate system, they can come to their own conclusions about global climate change based on the results that they generated.

When designed correctly, these tools can greatly enhance student learning and engagement in a classroom setting. Poor usability can frustrate users and discourage the adoption of technology into classrooms. In order to ensure a high level of usability, coproduction can be implemented to connect stakeholders to researchers and close the “usability gap” that exists between science makers and science users (Meadow et al., 2015). The main outcome of coproduction is usability, and this outcome should be tested in new ways that align more with its role in improving science and outcomes in the society (Bell et al., 2011; Meadow et al., 2015). The usability of coproduced tools can be measured by monitoring the performance of users on tasks with regard to the ease of use, the task time, and the user's perception of the experience of the software application or systems such as EzGCM (Pretorius et al., 2005).

Eye tracking can be very valuable in usability testing because it allows researchers to gain insight into user's eye movements while viewing a stimulus, like a website. When analyzing

users eye motions, data is generally split into fixations and saccades. A fixation is a moment where the eye is relatively motionless, and a saccade is a quick movement between fixations to another element and both are captured by the eye tracker (Ehmke, C., & Wilson, S., 2007). This data can be translated into gaze plots, which may be used to show the chronological succession of eye motions on a screen or webpage for an individual user, or heat maps, which display how long each part of a screen has been looked at. These visualizations of eye-tracking data can then be interpreted researchers to identify confusion or engagement on the part of the user, reading or scanning behaviors, or areas that are not garnering attention from users (Ehmke & Wilson, 2007). The results of usability testing are then communicated to stakeholders in order to improve the overall design of the website. In this study, eye-tracking will be used to understand how students navigate the EzGCM web tool with the goal of creating a framework of key design elements that creators can reference when making effective scientific teaching tools.

Visual Attention and Website Design

There are many website design recommendations from user experience experts that maximize usability, reduce usage time, eliminate frustration, and increase satisfaction and retention on a website (Tichindelean, et al. 2021). One example is page scanning directions where Pinto et al. (2013) and Connor et al. (2004) suggest that top-down scanning patterns are identified as preferable for goal-oriented task-solving web visualizations. Recently, Hung and Wang (2020) claimed that scan paths were predominantly vertical and horizontal in their study. Similarly, Buscher et al. (2009) considered that users typically look at the top left area of the screen in the first moments of engaging with the website. When five different layouts of websites were tested in a study by Tichindelean and colleagues (2020), they confirmed the previous findings and argued that the box-structured layout proved to be useful in guiding the gaze in a

vertical and horizontal scan path, rather than chaotic. In addition to website layout, elements like icons and symbols can aid in user attention. In an eye tracking study, Bucher and Schumacher (2006) confirmed this by contesting that colors and icons guide user attention and ensure that the user will interact with the intended website elements when hierarchical design elements are employed.

Eye Tracking for Usability

Eye-tracking has been utilized in previous studies to assess the usability of websites of tools designed for learning. For example, Zardari and colleagues (2021) used eye-tracking and post interviews to evaluate the usability of an e-learning portal known as QUEST. This research found that eye-tracking provided meaningful and relevant data that could be used to improve the experience of students using the website in an iterative process. Similarly, dynamic science tools, like mapping websites, have also been eye tracked to help researchers understand how users interact with the content (Manson et al., 2012; Alaçam & Dalc, 2009). Manson and colleagues (2012) found that when utilizing a mapping tool to complete tasks, many users preferred the familiar rectangle zoom or wheel zoom over the less intuitive and often touchy click zoom. Additionally, Alaçam and Dalc (2009), explore how users navigate iconic representation of data as it is displayed in popular mapping websites and found that icons can serve as an efficient way to communicate commands to users.

Finally, previous usability studies, like the one conducted by Wang and colleagues (2019) had the goal of aiding a redesign of an educational tool used to teach students math much like the current study seeks to inform a redesign of an educational climate modeling website. By giving students tasks to complete using the website and eye tracking them as they moved through the

tool , researchers were able to identify characteristics of the tool that helped students navigate the tasks as well as areas of potential confusion within the site (Wang et al., 2019).

RESEARCH QUESTIONS

The current research study aims to understand the most usable design features of the EzGCM educational website tool through testing two versions, A and B, with undergraduate students. Our research questions include:

1. What feature changes between the A and B version of EzGCM impact the accuracy and efficiency for users of the website?
2. What is the overall level of satisfaction by users of EzGCM?

MATERIALS AND METHODS

The goal of this research is to evaluate the usability of two versions of the EzGCM website through a mix of quantitative and qualitative methods.

Coproduction Process

The goal of this research is to evaluate the usability of two versions of the EzGCM website through a mix of quantitative and qualitative methods. To fulfill this goal, researchers and the stakeholder with EzGCM first worked together using a coproduction approach to develop a beta version of the software based on expert testimonial for the stakeholder as well as user design framework. The researcher and stakeholder met regularly and discussed possible areas of the website they may impact the user experience such as button location, button color, and labels. These discussions were used in conjunction with user design framework and existing research on user satisfaction to create a beta version (B) to test against the original version (A). The research is an example of co-production, where the stakeholder and researcher worked

together to create a more usable product for consumers to use in the classroom. The redesign of the EzGCM includes changing elements on the three main pages of the website: run simulations, post processing, and visualizations. Ongoing conversations between the stakeholder and researcher were held on a nearly bi-weekly basis to establish what changes and level of instruction would be utilized to create the (B) version of the software.

Development of A&B Designs

All of the knowledge on scan paths, the order in which a user navigates a page, were taken into consideration when designing the (B) version of EzGCM which is designed to have users begin the task in the top left of the screen in a box structure and follow a downward vertical motion in order to satisfy the task before moving horizontal. In addition, colors, contrasts, and changes in verbiage were added to the (B) version of the EzGCM website guided by previous research in visual attention. Because EzGCM utilizes a spatial visualization similar to those in Manson and colleagues' study, this information can serve as a framework for features that users may prefer in EzGCM.

For run simulations, abbreviated as RS, the original (A) and redesigned (B) interface is shown in figure 1A and figure 1B. On this page, changes were made to the location of the “pause” and “play” button moving it below the selection menu in the new version. Additionally, the proceed to post processing button was changed from grey to green and an iconic arrow was added to the button. Changes to button colors, contrast, and labels can influence click through and thus affect the efficiency of navigation, which can be tested through eye tracking (Spool et al., 1999; Bojko,2013). The redesigned version of the run simulations page can be seen in figure 1B.

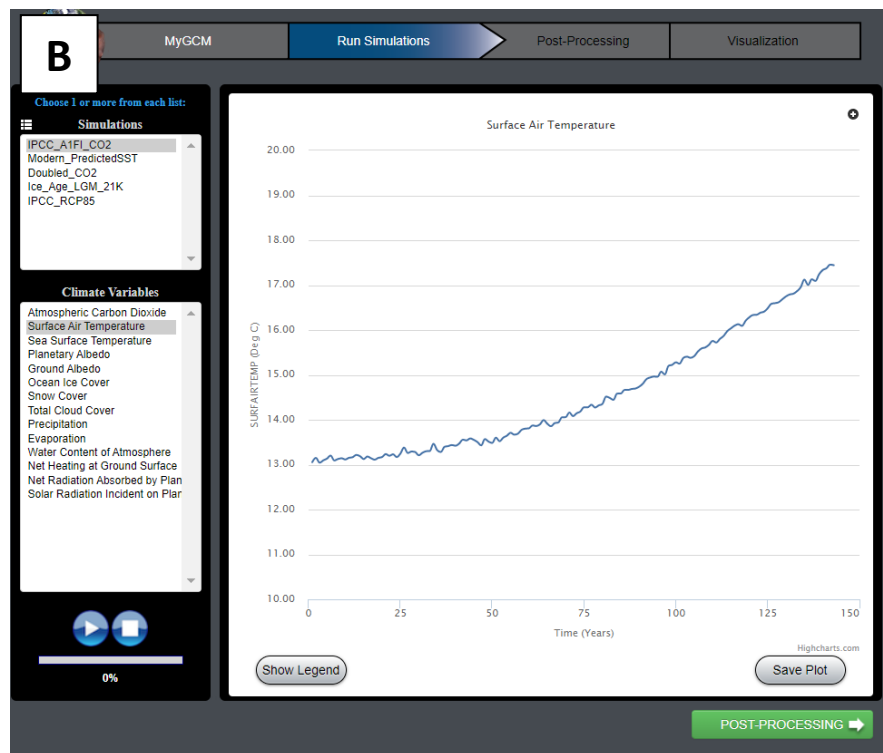
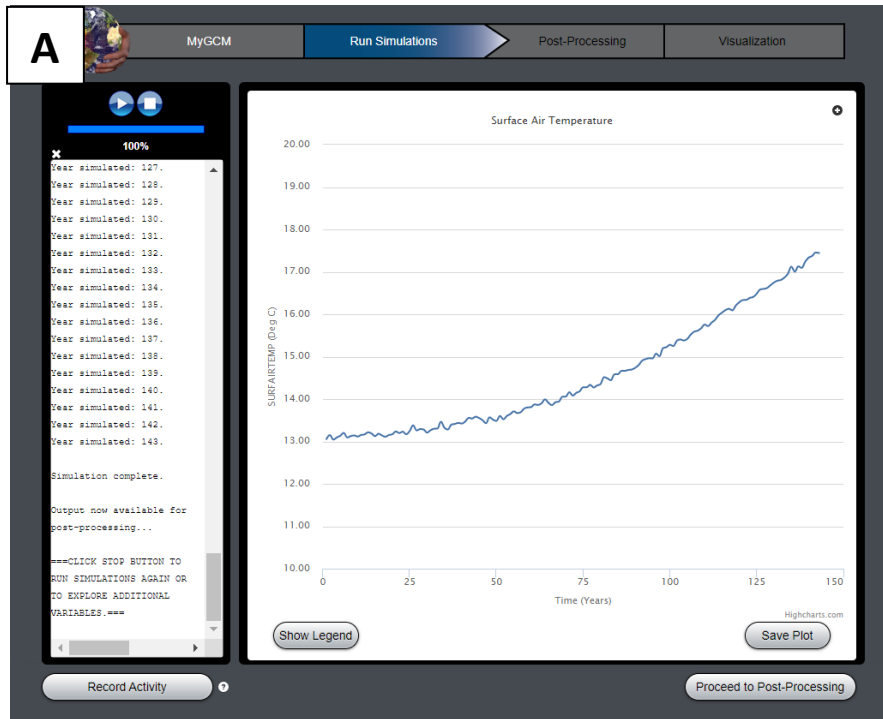


Figure 1: Interface display of version (A) (top) and version (B) (bottom) on the first step of the EzGCM webtool, the run simulations page.

A

MyGCM Run Simulations **Post-Processing** Visualization

Step 1. Average

Simulations

- IPCC_A1FI_CO2
- Modern_PredictedSST
- Doubled_CO2
- Ice_Age_LGM_21K
- Sturlian_cold_50m
- Varanger2_006_100pb

Years

- 1958
- 1959
- 1960
- 1961
- 1962
- 1963
- 1964
- 1965
- 1966
- 1967
- 1968
- 1969
- 1970
- 1971
- 1972
- 1973
- 1974
- 1975
- 1976
- 1977
- 1978
- 1979
- 1980

Average

Step 2. Extract

Existing Averages

- IPCC_A1FI_CO2@1958-1959
- Modern_PredictedSST@1958-1
- IPCC_A1FI_CO2@1958-1958
- IPCC_A1FI_CO2@1960-1960
- IPCC_A1FI_CO2@2096-2100

Time Periods

- Monthly
- Seasonal
- Annual

Climate Variables

Variable Name	Unit	Short Name
<input type="checkbox"/> Ocean Ice Coverage	Percent	OceanIce
<input type="checkbox"/> Snow Coverage	Percent	SnowCover
<input type="checkbox"/> Snow and Ice Coverage	Percent	SnowandIceCover
<input type="checkbox"/> Precipitation	mm/day	Precip
<input type="checkbox"/> Evaporation	mm/day	Evap
<input type="checkbox"/> Total Cloud Cover	Percent	TotCloudCover
<input type="checkbox"/> Planetary Albedo	Percent	PlanetAlbedo
<input type="checkbox"/> Soil Moisture	Percent	SoilMoist
<input type="checkbox"/> Surface Air Temperature	Deg C	SurfAirTemp
<input type="checkbox"/> Ocean Mixed-layer Temperature	Deg C	SST
<input type="checkbox"/> Snow Fall	kg/m**2	SnowFall
<input type="checkbox"/> Max Surface Air Temperature	K	MaxSurfAirTemp

Extract

Step 3. Visualization

Simulation Files

- Modern_PredictedSST.1958-1959
- IPCC_A1FI_CO2.1958-1959

Download Visualize

Record Activity Proceed to Visualization

B

MyGCM Run Simulations **Post-Processing** Visualization

Step 1. Average

Simulations

- IPCC_A1FI_CO2
- Modern_PredictedSST
- Doubled_CO2
- Ice_Age_LGM_21K
- IPCC_RCP85

Years

- 1958
- 1959
- 1960
- 1961
- 1962
- 1963
- 1964
- 1965
- 1966
- 1967
- 1968
- 1969
- 1970
- 1971
- 1972
- 1973
- 1974
- 1975
- 1976
- 1977
- 1978
- 1979
- 1980

Average

Step 2. Extract

Existing Averages

- Modern_PredictedSST@2099-2
- IPCC_A1FI_CO2@2099-2100

Time Periods

- Monthly
- Seasonal
- Annual

Climate Variables

Variable Name	Unit	Short Name
<input type="checkbox"/> Ocean Ice Coverage	Percent	OceanIce
<input type="checkbox"/> Snow Coverage	Percent	SnowCover
<input type="checkbox"/> Snow and Ice Coverage	Percent	SnowandIceCover
<input type="checkbox"/> Precipitation	mm/day	Precip
<input type="checkbox"/> Evaporation	mm/day	Evap
<input type="checkbox"/> Total Cloud Cover	Percent	TotCloudCover
<input type="checkbox"/> Planetary Albedo	Percent	PlanetAlbedo
<input type="checkbox"/> Soil Moisture	Percent	SoilMoist
<input checked="" type="checkbox"/> Surface Air Temperature	Deg C	SurfAirTemp
<input type="checkbox"/> Ocean Mixed-layer Temperature	Deg C	SST
<input type="checkbox"/> Snow Fall	kg/m**2	SnowFall
<input type="checkbox"/> Max Surface Air Temperature	K	MaxSurfAirTemp

Extract

Step 3. Visualization

Simulation Files

- IPCC_A1FI_CO2.2099-2100

Download

← RUN SIMULATIONS VISUALIZATION →

Figure 2: Interface display of version (A) (top) and version (B) (bottom) on the second step of the EzGCM webtool, the post processing page.

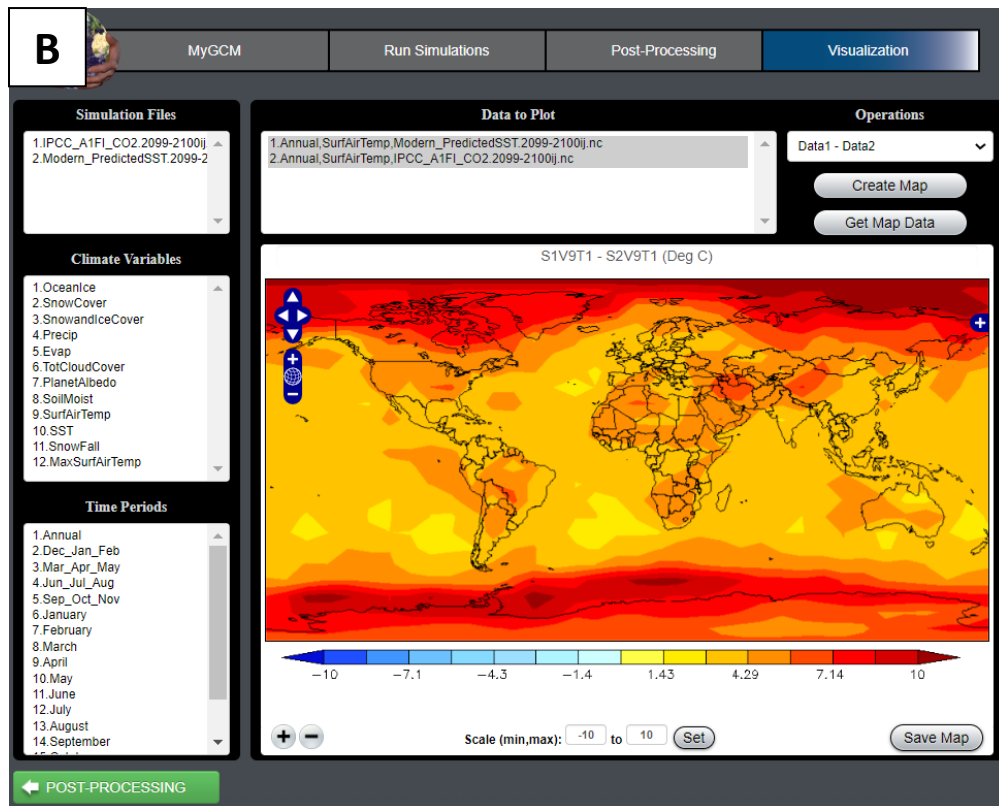
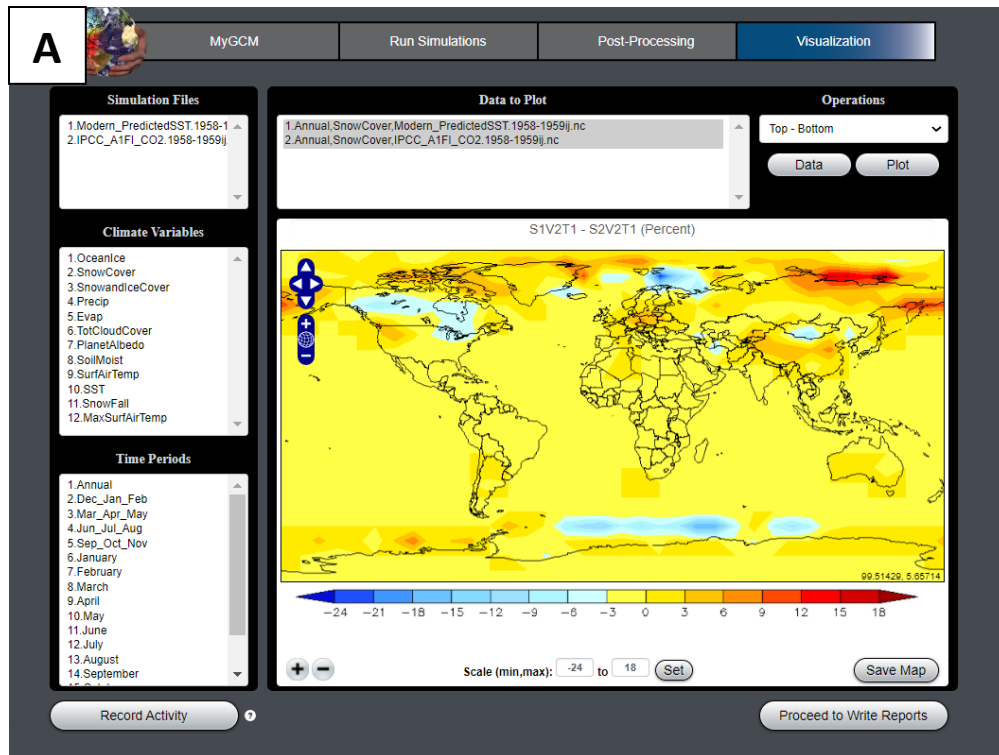


Figure 3: Interface display of version (A) (top) and version (B) (bottom) on the second step of the EzGCM webtool, the post processing page.

For the post processing page, abbreviated as PP, (Fig 2A & 2B), the detail of instructions for the same task were varied between treatment groups because the stakeholder wanted to test the influence of instruction on participants ability to navigate this page. On version (A), participants were given point and click instructions while users of version (B) were given more vague instructions with the same goal. The only change made to the interface of this page was the proceed to visualizations button which was kept the same as the previous page.

On the final page, visualizations, abbreviated as Vis, (Fig 3A & 3B), the labels for buttons were changed along with the operations menu. The operations menu is utilized in the task to perform subtraction of two files in order to create an anomaly map. In the (A) version, the operations dropdown menu had the first option of “top – bottom” and the second option of “bottom – top” referring to the files selected in the data to plot menu. Expert testimonial from educators using EzGCM in a classroom identified the operations menu as being a source of confusion for students who did not understand what the selection was referring to as top and bottom. The stakeholder expressed the desire to try out a new menu type, however, coding a dynamically populating menu is time consuming and challenging. A dynamically populated menu would mean that the options in the dropdown menu change depending on what is highlighted in the data to plot window. In this study, a dynamically populated menu was simulated menu without actually creating one for the purpose of this study to determine if this change would be worth devoting resources to in the future. To simulate this, the menu said “data1 – data2” and “data2 – data1” which more explicitly refers to the files created by the user in the task. A summary of these changes can be seen in table 1.

Table 1: Outline of the design changes and justification for each page of EzGCM.

Page	Design Change	Justification
Run Simulations	Play and pause buttons moved to under the selection menus on left most pane.	(Buscher et al, 2009).
Post processing	The level of detail in instruction will vary between users in treatment group A and B.	Stakeholder request for instructional purposes
Visualizations	Operations menu changed from “top – bottom” to a simulated dynamically populated menu that says “data x – data y”	Stakeholder request based on user testimonial.
	“Plot” changed to “create map” and “data” changed to “Get map data”	(Cox, 2002)
All Pages	Button to advance pages is changed from grey to green and symbology of arrow added.	(Richardson et al., 2009). (Johnson, 2020)

Participant Task

In addition to designing the new version of the webtool, the researcher and stakeholder met and developed a task that would allow participants to interact with all the design changes and be typical of something performed by the intended user of the website. While completing the task, the participants were also instructed to answer embedded questions that related to successful task completion. For the task, participants first began with the run simulations page, where they were instructed to select the modern predicted climate simulation from the simulations menu, then the surface air temperature variable from the climate variables list and run the simulation using the pause/play button. After the simulation was complete, they repeated the previous steps adding the IPCC RCP85 simulation. On this page, they were asked to relay the

global average temperature from each simulation as well as save their plot as a measure of task completion. On the subsequent post processing page, participants were instructed to average the last two years of each simulation and extract surface air temperature as well as two other climate variables. Finally, on the visualizations page, they first were instructed to prepare the files for visualization by selecting each simulation, surface air temperature, and the annual temporal range from the menus. Then, participants created two maps of surface air temperature using each simulation, making use of the plot or create map button. After, they had to use the operations menu to subtract the two simulations and create their resulting map. From this, they were asked to read the temperature value for Alabama, an area that is familiar to participants due to the study location, scale the map, and relay the temperature value again for the same location. Finally, participants were instructed to save their map which was used as a metric for accuracy (see appendix, table 2, for the exact task instructions).

Pilot and Recruitment

Before recruitment of participants began, three volunteers participated in a pilot study to refine details of the study procedure as well as identify if any obvious changes needed to be made to EzGCM or the task instructions. One slight error in design was identified on the run simulations page of the redesigned version of the webtool that needed refinement from developers before proceeding with the data collection for that version. However, data collection proceeded with the original version of EzGCM while developers fixed the error in order to expedite the data collection process. The data from the pilot study was not used for analysis and was strictly used for ensuring that the website and instructions were suitable before usability testing began.

Researchers recruited undergraduate participants from a large University in the southeastern U.S. from introductory STEM and non-STEM classes. In total, 23 participants were given the redesigned version and 28 students were given the original version of EzGCM. Recruitment began by going into undergraduate classrooms across multiple disciplines and explaining the study to students, putting up flyers across campus, as well as sending out a volunteer email to office administrators in each department and asking those staff members to forward the email to their undergraduate students. The sample in this study is considered a convenience sample due to the manner of recruitment coming from familiar classes, however, it is still a random sample due to voluntary participation as well as random assignment to website groups. There was no requirement of prior knowledge about climate change as well as no requirements of certain demographics for participants. Undergraduate participants were asked to schedule an appointment to come and complete the research which took place in a Geocognition Lab at a large southeastern university.

Intervention with Eye Tracking

Each appointment lasted approximately 60 minutes. Participants were first given time to review and sign the informed consent document that was approved by the Institutional Review Board (IRB) and verbal confirmation of their voluntary participation in the study was obtained by the researcher. Then, participants were instructed to come up with a four-digit code that was used on each step of the data collection in order to link their pre survey, post survey, eye tracking data, and interview while ensuring anonymity and compliance with the IRB requirements. Participants were then given a pre-survey which included some demographic information, background and familiarity with climate modeling, a few Likert style questions where they ranked their self-efficacy with webtools and computer skills, and a few climate specific content

questions. After the pre-survey was completed, participants were assigned either the (A) or (B) study group, users completed the same tasks regardless of treatment group.

Eye tracking technology was used to test EzGCM version (A) & (B), each with intentional design variations that may influence the efficiency, accuracy, and satisfaction of users as they use the site to complete a series of tasks. Tobii eye-tracking software was used as a means of measurement with the EzGCM website interface. Tobii Pro software was used to design and set up the experiment and measurements were recorded with the Tobii TX300 eye-tracker. Before the eye-tracking portion of the study began, the researcher calibrated participants' eyes with the eye tracker to establish accuracy of the collected results. The calibration process required that the participant follow red dots across the screen and calibrations of >50% weighted gaze samples were allowed to continue to the main study. The Tobii TX300 is a 300 Hz desktop eye-tracking instrument attached to the bottom of a computer monitor. A 300 Hz instrument like the one utilized in this study collects 300 data points per minute which is much higher resolution than instruments typically used in usability studies. In fact, usability research requires a much lower sampling rate than what the instrument used in this study has measured, thus, there is some flexibility in the generally accepted threshold (Bucher & Schumacher, 2006). This remote eye-tracker is hands free, requires participants to sit in one place, and it is noninvasive compared to traditional eye trackers (Bojko, 2013).

After successful calibration, participants were eye tracked as they navigated their assigned version of EzGCM and completed assigned tasks as well as answered embedded questions about the tasks. If users were stuck at a step and asked for help, they were instructed by the researcher to explore the page to find the correct answer, and this was noted in the

researcher's lab notebook. The researcher also observed participants and recorded qualitative notes.

Post Survey and Interview

After the eye tracking session, participants completed a user experience post-survey to rank their satisfaction and ease of use for the product as well as a semi structured retrospective interview. The post survey consisted of 5 Likert style questions to gauge participants feelings about the webtool. During these interviews, which lasted approximately 10 minutes per participant, the researcher played the users gaze replay and asked them questions to allow them the opportunity to elaborate on their thoughts. A gaze replay is "...video replay of the task with eye-movements collected during the tasks superimposed over the stimulus" (Bojko 2013). This follows the retrospective verbal protocol that includes playing some type of visual (gaze replay) and have the participant verbally recall what they were thinking throughout the task (Cho et al. 2019; Bojko 2013). In usability research, the retrospective interview is often utilized rather than other methodologies, such as a think aloud interview, so as not to distract the participant and interfere with the data collection from the eye tracking instrument. Interviews and satisfaction survey results were used to triangulate and support data collected from quantitative eye tracking. This methodology aligns with the sequential explanatory strategy indicative of a mixed methods design which involves "...the collection and analysis of quantitative data in a first phase of research followed by the collection of and analysis of qualitative data in a second phase that builds up on the results of the initial quantitative results" (Creswell, 2009). This design strategy is used to explain the quantitative data (e.g., eye-tracking) by collecting the qualitative data (interviews). Interviews are helpful for data collection because eye-tracking by itself cannot help

with answering questions about people’s actions (Bojko, 2013). Post survey questions, interview protocol, and sample interview questions are located in the appendices section.

Table 2: Validity and reliability with corresponding study characteristics.

Criteria	Study Characteristics
Validity	Pilot study Eye tracking calibration Co-produced research design Expert review of task assessment Research-grade survey instruments
Reliability	Triangulation through mixed methods Robust sample size Research-grade survey instruments Multiple item task assessment

ANALYSIS

The eye-tracking experimental design included measurements for users’ navigation of the website in regard to efficiency, effectiveness, and satisfaction with the resource. The measures, source information and analysis methods are outlined in table 2.

Table 3: Objectives and means table for usability of EzGCM.

Objectives	Candidate Measures	Source of Information	Analysis
Usability of versions A and B	Efficiency (How fast can tasks be completes)	Eye tracking data. specifically, time to first click and fixation duration in various areas of interest.	t-tests between versions. ANOVA within versions using pre-survey questions.
	Accuracy (How accurate the tasks are completed)	Embedded Q&A correctness, screenshots of interface where it applies.	t-tests between (A) and (B) versions.
	Satisfaction (How do users feel after completing the task)	Post survey and interview data.	t-test between groups for post satisfaction survey. Example excerpts and researcher themes.

Data Preparation

Before beginning analysis, multiple adjustments had to be made to the bulk eye tracking and survey data in order to be compatible with the analysis software.

Areas of Interest

In eye tracking analysis, areas of interest (AOIs) are boundaries that a researcher draws on the playback of a participant’s eye tracking footage to limit the analysis to one or several defined areas. AOI’s are common in quantitative eye tracking analysis to define areas interesting for the research (Bojko, 2013), for example, part of the display on a product or a navigation bar on a web page.

In the case of this study, AOIs were drawn around all of the changes made to the website including the pause and play button, the proceed to post processing button, the instructions on the post processing page, the proceed to visualizations button, the operations menu, and the create map or plot. In addition, AOIs were drawn around places that may be impacted by

changes made to the website or be required to complete the task such as the instructions on the run simulations page, the simulations menu on the run simulations page, the post processing page selection area, the instructions on the visualizations page, the map area on the visualizations page, and the button that adds or removes a map button on the visualizations page. A visual representation of these AOIs for each page of the webtool can be seen in figure 4, figure 5, and figure 6. When conducting eye tracking a static product, AOIs can be imported from one participant and generalized to the rest, however, for more interactive websites, the data for each participant must be mapped individually. When AOIs overlap, meaning that during one time in the session an AOI would be present on the screen and at another time that a different AOI would fully or partially cover the original AOI, the original AOI must be deactivated and the new AOI activated during this time. Due to the highly dynamic nature of the website with some areas overlapping others, and the users completing the task at different cadences, eye tracking playback had to be played by the researcher and AOIs had to be manually drawn for each participant. Throughout this process, the playback had to be paused and AOIs had to be activated or deactivated depending on what stimuli was present on screen at that time. One case of an overlapping AOI can be seen through the operations menu and plot (A) or create map (B) button. When the operations menu is closed, both AOIs can be present on screen, however, when the user opened the operations menu, the plot or create map was covered up and had to be deactivated for the time that is it not shown to the user this example can be seen in the inlay of figure 6. The process of drawing AOIs manually required about 30-45 minutes per participant depending on their level of interaction with the tool and the length of their eye tracking session.

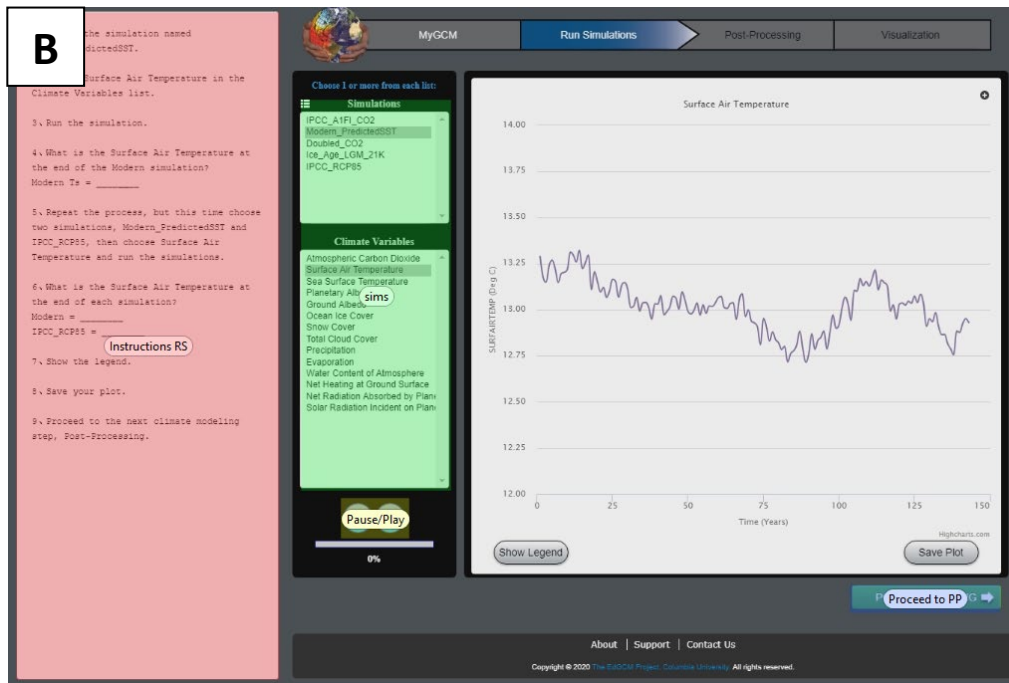
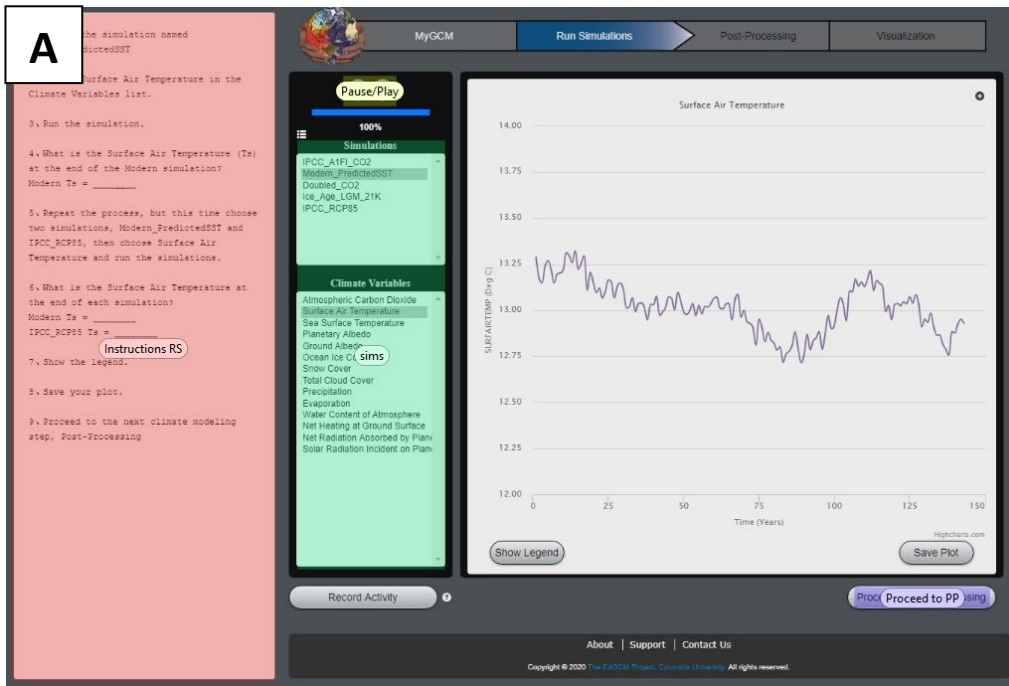


Figure 4: Areas of interest for the run simulations page version (A) and version (B).

MyGCM
Run Simulations
Post-Processing
Visualization

A

the simulation
dictatedSST from the Simulations
Step 1 column.

2. Select the Last 2 years from the Years list.
3. Click the Average button.
4. When the processing is complete click the Close button. Then redo steps 1-3 using the simulation IPCC_RCP85.
5. When both simulations have been averaged go to the Existing Averages list in the Step 2 Extract column and select both the Modern and IPCC datasets.
6. Select Surface Air Temperature and at least 2 more variables from the Climate Variables list. [Instructions PP](#)
7. Click the Extract button.
8. When the extraction process is complete click the Close button. Then select the Modern and IPCC_RCP85 datasets from the Visualization list in the Step 3 Visualization column.
9. Click on the Proceed to Visualization Button.

Step 1. Average

Simulations

IPCC_A1FI_CO2
Modern_PredictedSST
Doubled_CO2
Ice_Age_LGM_21K
IPCC_RCP85

Years

1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873

Next 2
Next 5
Last 2
Last 5

Average

Step 2. Extract

Existing Averages

Modern_PredictedSST@2099-2100
IPCC_RCP85@2099-2100

Time Periods

Monthly
 Seasonal
 Annual

Climate Variables

Variable Name	Unit	Short Name
<input type="checkbox"/> Ocean Ice Coverage	Percent	OceanIce
<input type="checkbox"/> Snow Coverage	Percent	SnowCover
<input type="checkbox"/> Snow and Ice Coverage	Percent	SnowandIceCover
<input type="checkbox"/> Precipitation	mm/day	Precip
<input type="checkbox"/> Evaporation	mm/day	Evap
<input type="checkbox"/> Total Cloud Cover	Percent	TotCloudCover
<input type="checkbox"/> Planetary Albedo	Percent	PlanetAlbedo
<input type="checkbox"/> Soil Moisture	Percent	SoilMoist
<input checked="" type="checkbox"/> Surface Air Temperature	Deg C	SurfAirTemp
<input checked="" type="checkbox"/> Ocean Mixed-layer Temperature	Deg C	SST
<input type="checkbox"/> Snow Fall	kg/m**2	SnowFall
<input checked="" type="checkbox"/> Max Surface Air Temperature	K	MaxSurfAirTemp

Extract

Step 3. Visualization

Simulation Files

Modern_PredictedSST.2099-2100.nc
IPCC_RCP85.2099-2100.nc

Download Visualize

Record Activity
Proceed to vis on

About | Support | Contact Us
 Copyright © 2020 The EaGCM Project, Columbia University. All rights reserved.

MyGCM
Run Simulations
Post-Processing
Visualization

B

the simulation
dictatedSST from the Simulations
Step 1 column.

2. Select years 2099-2100 from the Years list.
3. Average the last 2 years.
4. When the averaging is complete close the window and redo steps 1-3 using the simulation IPCC_RCP85.
5. When both simulations have been averaged select both the Modern and IPCC files from the Existing Averages list in the Extract column.
6. Select Surface Air Temperature and at least 2 more climate variables. [Instructions PP](#)
7. Extract the data.
8. When extracting is complete close the window and select the Modern and IPCC_RCP85 files from the Visualization list.
9. Proceed to Visualization.

Step 1. Average

Simulations

IPCC_A1FI_CO2
Modern_PredictedSST
Doubled_CO2
Ice_Age_LGM_21K
IPCC_RCP85

Years

2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100

Next 2
Next 5
Last 2
Last 5

Average

Step 2. Extract

Existing Averages

Modern_PredictedSST@2099-2100
IPCC_RCP85@2099-2100

Time Periods

Monthly
 Seasonal
 Annual

Climate Variables

Variable Name	Unit	Short Name
<input type="checkbox"/> Ocean Ice Coverage	Percent	OceanIce
<input checked="" type="checkbox"/> Snow Coverage	Percent	SnowCover
<input type="checkbox"/> Snow and Ice Coverage	Percent	SnowandIceCover
<input type="checkbox"/> Precipitation	mm/day	Precip
<input type="checkbox"/> Evaporation	mm/day	Evap
<input type="checkbox"/> Total Cloud Cover	Percent	TotCloudCover
<input type="checkbox"/> Planetary Albedo	Percent	PlanetAlbedo
<input type="checkbox"/> Soil Moisture	Percent	SoilMoist
<input checked="" type="checkbox"/> Surface Air Temperature	Deg C	SurfAirTemp
<input type="checkbox"/> Ocean Mixed-layer Temperature	Deg C	SST
<input type="checkbox"/> Snow Fall	kg/m**2	SnowFall
<input type="checkbox"/> Max Surface Air Temperature	K	MaxSurfAirTemp

Extract

Step 3. Visualization

Simulation Files

Modern_PredictedSST.2099-2100.nc
IPCC_RCP85.2099-2100.nc

Download

← RUN SIMULATIONS
Proceed to vis →

About | Support | Contact Us
 Copyright © 2020 The EaGCM Project, Columbia University. All rights reserved.

Figure 5: Areas of interest for the post processing page version (A) and version (B).

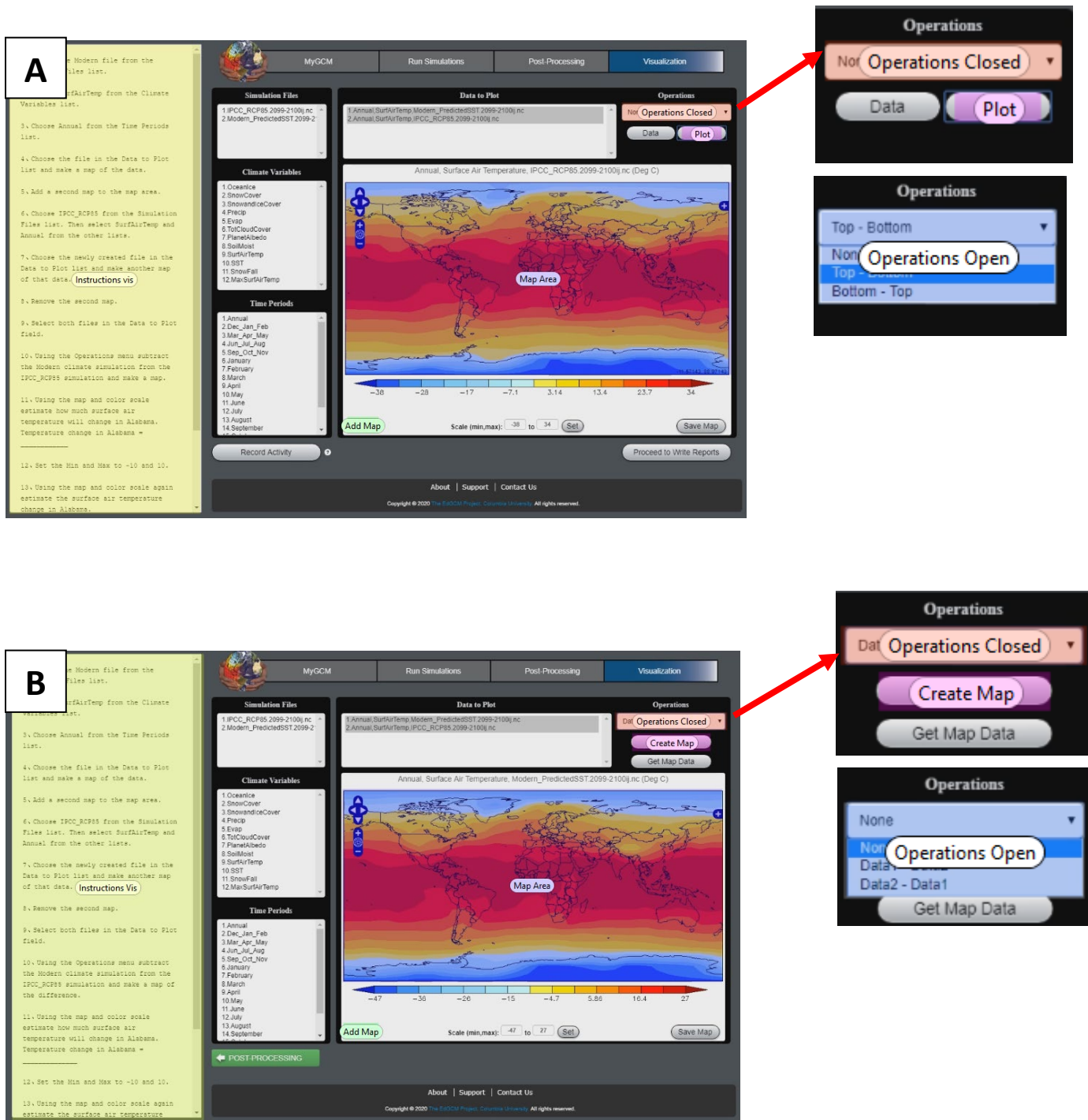


Figure 6: Areas of interest for the visualizations page version (A) and version (B). The inlays show an example of an overlapping AOI for each version.

Excel Preparation

After the AOIs for every participant were drawn, the eye tracking metrics were exported to an Excel file. The eye tracking metrics used in this study were total fixation duration and time to first click. Fixation duration is the time that the eye remains still when focusing on some stimuli (Holmqvist et al. 2011). Fixation duration in an AOI differs from the total amount of time spent in an AOI because during a fixation, the eye is motionless which means a user is taking in information rather than moving from one location to another, which is known as a saccade (Bojko, 2013). Time to first click records a timestamp for a user's clicks on a particular stimuli on the website which are contained within an AOI. Time to first click is a good metric for efficiency when the task requires the participant to interact with the AOI by clicking.

Because the participants' AOIs were mapped individually, they had to be aggregated in Excel so that each AOI was contained within a column and each participant was contained within a row. After creating the proper format, the data had to be cleaned prior to analysis. As part of this process, a duplicate of the data was created so that it could be manipulated. Participants with a gaze weight of less than 50% were deleted from the spreadsheet with data about total fixation duration, however, they were maintained in the sheet containing the information about time to first click as a low gaze weight does not impact the eye tracking instruments ability to record mouse clicks. After, zeros were replaced with blank values for both metrics as zeros represent a participant who did not fixate in or click in a particular AOI. It is important to remove the zeros because in the analysis, the researchers wanted to focus on participants who did fixate or click with an AOI and needed to omit those who did not from the analysis so that the data would not be negatively skewed. After this, each sheet was divided into two that contained participants who received the original, A, version of the website in one sheet

and those who received the redesigned, B, version of the website in a separate sheet. The last transformation that needed to occur was to normalize the time to first click data in order to get a realistic understanding of the trends occurring on each page. Because each page has a specific task with instructions, AOIs on the second page, post processing, and AOIs the third page, visualizations, were normalized to only include the time spent on that page. By normalizing the data per page for each participant, it ensures that the performance on one task does not influence the subsequent tasks. Normalization was performed by subtracting the time spent on the previous page from the time recorded for each AOI to get the time to first click in the AOI from the start of that page rather than from the start of the session. Total fixation duration did not need to be normalized because that measurement is dependent on the duration spent by users in a particular AOI rather than duration of the entire session. The final four sheets were saved as .csv files with appropriate names. All data analysis took place in R studio.

Quantitative and Statistical Analysis

All data analysis took place in R studio. T tests for samples with unequal variances were used to compare the data collected between version groups. Researchers considered the assumptions for a standard t test prior to running them and concluded that this was a suitable test to analyze the data because the samples met the requirements for independence, and they were randomly sampled. In addition, Q-Q Plots were created for each dataset to check normality, and all were fairly normal with few outliers outside of the accepted buffer zone. Finally, the variances were not equal between groups, however, outliers of participants with low gaze weight were deleted to make the variances more similar and the selected t test was for samples with unequal variances to account for this.

Efficiency

Efficiency, which is a measure of how fast a user can complete a task or accomplish a goal, was measured using quantitative eye tracking data. After the eye tracking test is performed the data for each participant is compiled in Tobii Pro. The start and stop for each page, run simulations, post processing, and visualizations, for each participant was logged using event markers in Pro Lab. The event markers are then used to define time intervals for each stage in the task for each participant. T-tests were run for applicable AOIs to compare the averages between versions of the website and determine if there was a significant difference between the two for each eye tracking metrics.

Accuracy

The methodology presented by Alaçam and Dalc, who gave the tasks in their study completion scores in order to assess accuracy can be applied to the current study on EzGCM (2009). Effectiveness, or accuracy, was measured based off of whether the participant chose the correct answer on the task and the saved plot and map was correct. All questions and products were scored in a binary fashion in which the number 1 represented a correct response or product, and the number 0 represented an incorrect product. Participants were required to answer a question on the run simulations page that involved them reading the end of the simulation on a plotted graph. If students plotted the incorrect simulations or variable, they would be unable to accurately answer the questions. In addition, participants may have gotten the answer incorrect if they interpreted the plot incorrectly. Participants were also instructed to save their plot on this page and the plots were also scored for accuracy. An example of a correct and incorrect plot from the run simulations page is shown in figure 7.

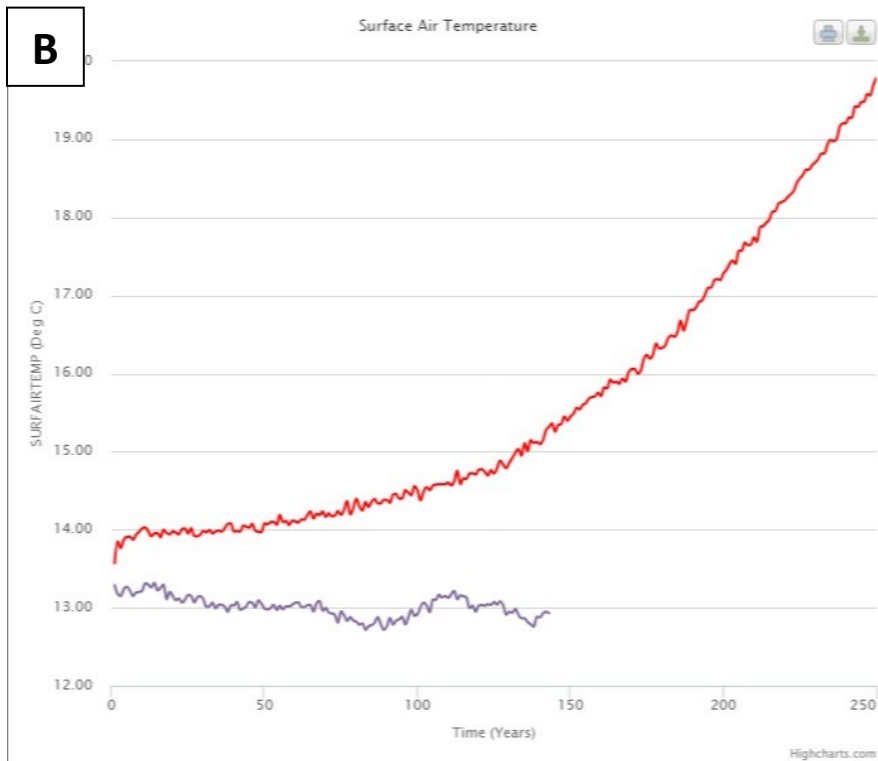
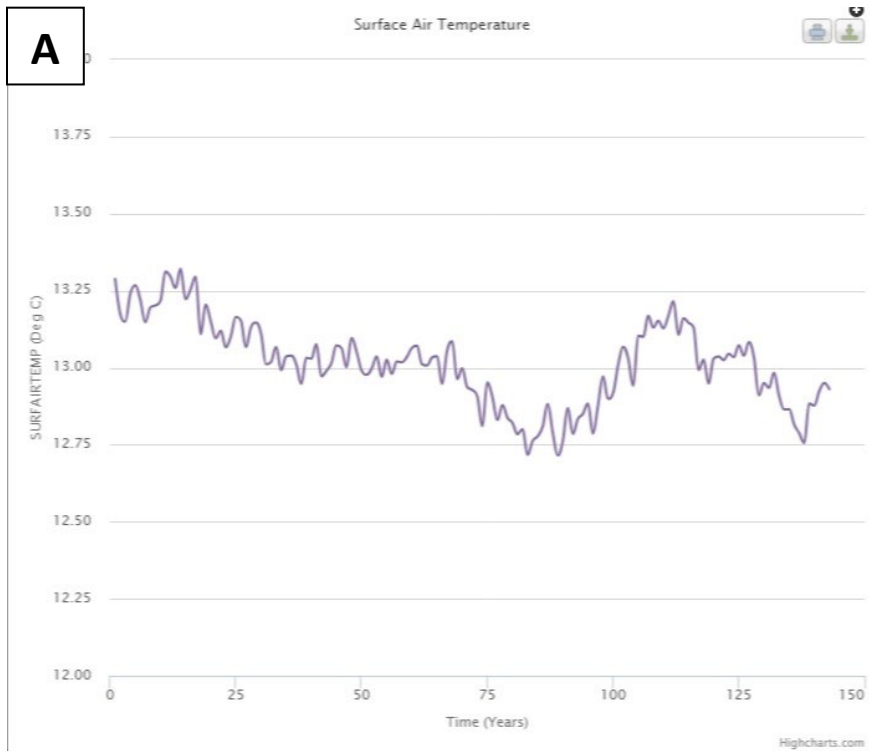


Figure 7: (A) correct participant example and (B) incorrect dummy example of saved plots.

Later in the session, participants were asked to create a map using the files that they had prepared on the previous pages. After creating and scaling the map according to the instructions, they were asked to relay the change in surface air temperature for Auburn, Alabama and save their final map. The answer to the question as well as the map was scored by researchers for accuracy. The most common reason a map was produced incorrectly was if the student performed the incorrect operation, which would display negative anomaly values. In addition, an incorrect map could also result from a user selecting the incorrect timeframe, which would change the intensity of the anomaly values, selecting the incorrect variable, which would display as a different geospatial distribution, or scaling the map incorrectly, which would be evident in the scale bar. The answer to this question as well as the resulting maps were scored for accuracy. An example of a correct map and a couple examples of incorrect maps is shown in figure 8. For analysis purposes, correct answers and graphics were marked as 1 and incorrect answers and graphics were marked as 0. Although all participants answered the questions embedded in the task, only 24 and 23 participants in the (A) and (B) version, respectively saved their plots from run simulations and 22 and 20 participants in the (A) and (B) version, respectively saved their map from the visualizations page. Analysis was only performed on saved graphics, meaning that if a student did not save a graphic, that participant was omitted from the analysis.

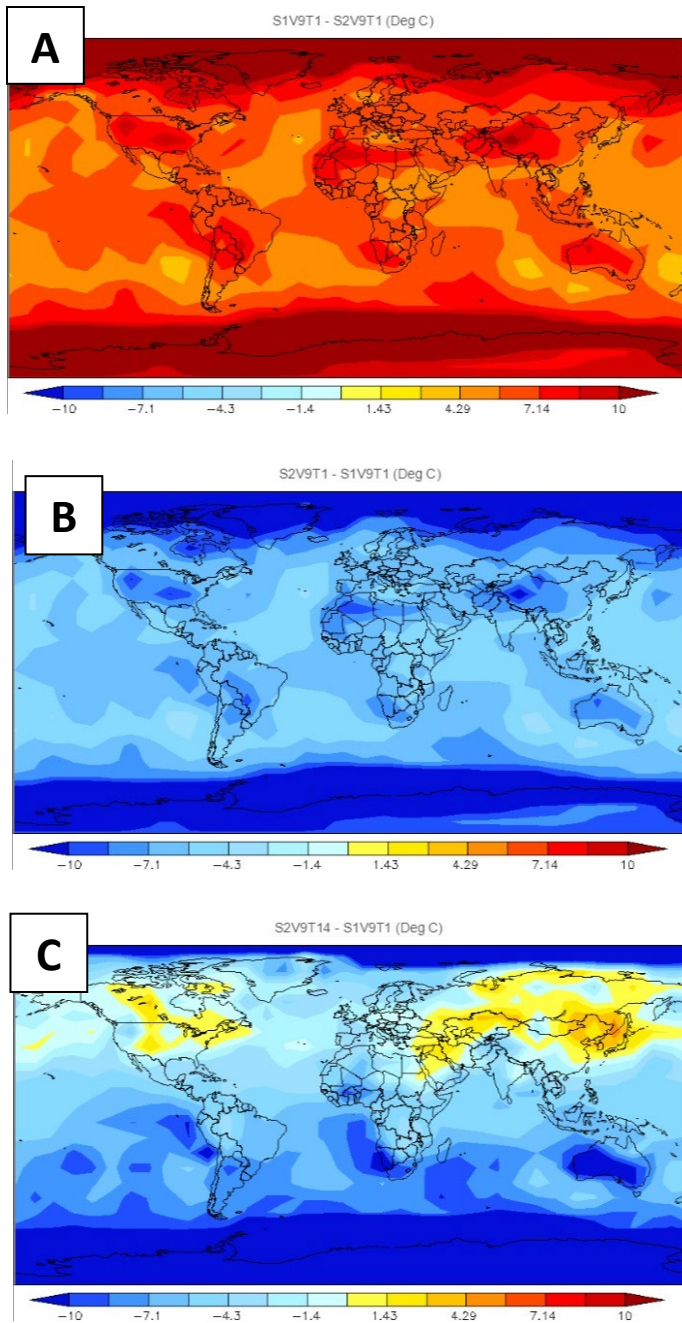


Figure 8: examples of correct (A) and incorrect (B, C) saved maps. A is correct because it shows the surface air temperature anomaly as forecasted by the IPCC in years 2099-2100 as instructed by the task. Additionally, this map is scaled correctly. B is incorrect due to the participant selecting the incorrect operation, while C is incorrect due to the incorrect time frame and operation being selected.

Satisfaction

Satisfaction, a measure of how users feel about the website, was evaluated qualitatively by the overall perceptions of EzGCM from retrospective interviews (Maudlin et al., 2020). Transcribed interviews were read by researchers and participant quotes that provided additional explanations for findings from the eye tracking analysis were isolated.

To analyze the post survey, the first four Likert style questions were translated into numerical scores (1-5) with 1 representing extremely unlikely/disagree, 2 representing unlikely/somewhat disagree, 3 representing neutral/ neither agree or disagree, 4 representing likely/somewhat agree, and 5 representing extremely likely /agree. Due to this coding scheme, values above 3 are seen as generally positive while values below 3 are seen as generally negative. The fifth question was a yes or no question where a was coded as a 1 and yes was coded as a 2.

Qualitative Analysis

Qualitative analysis was utilized on participant interviews and heatmaps and gaze plots created from eye tracking data. The rich qualitative data collected in this study was used to triangulate and explain results that were obtained quantitatively.

Interviews

Once all retrospective interviews were complete, the audio was transcribed by a transcription service. After all transcriptions were obtained, the files for each version group were aggregated into a single document. The researcher looked through each document and isolated quotes that discussed the specific design changes made between versions. These quotes and the accompanying themes are located in table 8.

Heat Maps and Gaze Plots

Heat maps are a common visualization of eye tracking data that provide a view of fixation density for a task overlaid on a static image (Bojko, 2013). Another common data visualization that is used to display eye tracking data is a gaze plot. Gaze plots, also called scan path, is a visual representation of one participant's visual path also presented on a static image (Bojko, 2013; Holmqvist et al, 2011). Each circle in the gaze plot represents a fixation and the size of the circle reflect the fixation duration, the lines linking the fixations are saccades. Since this type of visualization can contain lots of fixation points, each fixation is numbered in chronological order so the path can be followed. Heat maps and gaze plots can be used to make qualitative deductions about overall patterns of users. In the current study, heat maps were created using the aggregated fixations per page for all participants for each page (figure 9, figure 10, and figure 11). Similarly, gaze plots were created to compare versions between groups from a single participant that spent approximately the same amount of time on each page (figure 12 and figure 13). These visualizations were then observed by the researcher and general trends were noted for each task and page to compare between groups.

RESULTS

Quantitative results

Efficiency

After omitting the 7 participants with a gaze weight less than 50%, there were 24 participants assigned version (A) and 22 participants in group (B). To analyze total fixation duration, researchers performed separate t-tests to compare each of the AOIs, of which the full results are shown in table 2 in the appendix. The analyses revealed that three AOIs, the

simulations menu (simulations), the proceed to post processing button (proceed to PP), and the plot or create button (plot/create) had a significant difference in average fixation duration between groups at the .05 significance level ($p=0.0065$, $p=0.0029$, $p=0.0057$, respectively). At the .1 significance level, the AOI surrounding the proceed to visualizations button (proceed to vis) also becomes significant ($p=0.0704$). For simulations, for proceed to PP and proceed to vis, the average fixation duration was greater for those who received the (A) version in each AOI. Contrastingly, the average fixation duration for those who received the (A) version of the website with the plot button was less than those who received the (B) version with the create map button.

Efficiency was also measured using time to first click between versions for applicable AOIs. For the metric of time to first click, all participant data was maintained which resulted in 28 who received version (A) and 23 who received version (B) of the website. Because participants were not instructed to click in all AOIs to complete the task, all AOIs were not important to compare between groups. This was especially evident in the AOIs surrounding the instructions, for example, because participants did not need to click the instructions to complete the task and thus, this comparison was irrelevant. Because of this, researchers completed t-tests for 9 AOIs, omitting the three AOIs that contained instructions (table 2, appendix). Of these, 2 AOIs, the one containing the operations menu while it was open, and the one containing either the plot or create map button were found to have significant differences in time to first click between groups ($p=0.0439$ and $p=0.0406$, respectively). Both of these AOIs are located on the visualizations page. The full statistical output for significant AOIs is shown in table 4. In both of these AOIs, the average time spent before the first click was longer for users given the (A) version of the website than it was for those given the (B) version.

Table 4: Statistical output for significant AOIs compared between groups. An asterisk denotes AOIs significant at the .05 confidence level.

Metric	Area of Interest	Avg A	Std dev A	Avg B	Std dev B	p-value
Total	Simulations	43.49	14.60	32.98	9.94	0.0065*
Fixation Duration	Proceed to PP	3.17	2.49	1.34	0.98	0.0029*
	Plot/Create	1.50	1.11	4.88	4.88	0.0057*
Time to First Click	Operations	207.89	116.41	94.66	207.10	0.0439*
	Plot/Create	96.17	100.39	46.47	30.65	0.0406*

Accuracy

When comparing the accuracy of the completed questions between groups it was found that 100% of participants in both groups got the question and saved plot on the run simulations page correct (table 5). Because of this, there was no need to run a t-test to compare the accuracy between groups for these tasks. However, on the visualizations page, the initial analysis revealed that 44.44% of participants in version (A) compared to 78.26% of the participants that received version (B) got the question on the visualizations page correct. Of the participants who saved their resulting maps, 54.54% in the (A) version group had a correct product and 76.19% of participants in the version B group had a correct product.

Table 5: Distribution of participant counts and correct responses between groups.

Question	A version		B version	
	Response count	% Correct	Response count	% Correct
Question RS	27	100	23	100
Question Vis	27	44.44	23	78.26
Plot RS	24	100	20	100
Map Vis	22	54.54	20	76.19

T-tests for the question asked on the visualization page yielded that there was a significant difference between version groups in accuracy with the (B) version having a greater average accuracy rate than the (A) version (Avg A=0.428, avg B=0.783, p=0.008749) (table 6). For the saved maps, the (B) group had a greater average accuracy rate than the B group, however, this difference was not significant at the .05 significance level (Avg A=0.545, avg B=0.762, p=0.1419).

Table 6: between group t-test results for the visualizations page. Asterisks indicate significance.

Question	Avg A	Std dev A	Avg B	Std dev B	t	Df	p-value
Question Vis	0.428	0.503	0.783	0.422	2.731	48.975	0.008749*
Map Vis	0.545	0.510	0.762	0.436	-1.498	40.538	0.1419

Satisfaction

T-tests were performed for each question to compare the results between groups (table 7). For all questions except question 2, which asked if their knowledge of climate change improved after using the tool, the average score for the (B) version was more positive than version (A). It is important to note, however, that question 2 has the smallest difference between the group averages. For question 3, which asked if the user felt like those with no prior experience with

climate data could use the website, the t-test revealed that there is evidence to suggest that there is a statistically significant difference between group responses with those who received the (B) version more likely to agree ($p= 0.01264$, Avg A= 3.571, Avg B= 4.348). Question 3 also had the largest difference between group averages out of all questions asked on the survey. T-test for question 5 revealed a significant difference in users agreeance with the statement that their perception of climate change differs from their pre-survey after completing the website task with the B version having a greater group average than the (A) version. ($p=0.0469$, avg A=1.571, avg B=1.826). Questions 1, 2, and 4 were not found to have statistically significant different responses between groups at the .05 significance level, however, at a slightly less conservative significance level of .1, question 4, which asks about the ease of navigation, becomes significant (avg a=3.5, avg b=4, $p=0.09181$).

Table 7: between group t-test results for post intervention satisfaction survey. Asterisks indicate significance.

Question	Avg A	Std dev A	Avg B	Std dev B	Avg difference	t	df	pvalue
Q1	3.392	0.875	3.739	1.010	0.346273	-1.293	43.903	0.2027
Q2	3.571	0.920	3.435	1.080	-0.13665	0.480	43.468	0.6334
Q3	3.571	1.136	4.348	1.027	0.776397	-2.559	48.514	0.01364*
Q4	3.5	1.072	4	1	0.5	-1.720	48.161	0.09181
Q5	1.571	0.504	1.826	0.388	0.254658	-2.039	48.818	0.0469*

Qualitative results

Transcribed interviews and eye tracking visualizations were qualitatively analyzed by the researcher who looked for trends and patterns within the data.

Interviews

Transcribed interviews from participants in each treatment group were read by the researcher and the pages were isolated as themes. Quotes were isolated that related to changes made on that page for each group (table 8).

Table 8: Participant quotes isolated between versions for each page.

Version	Theme	Example Quote
A	Run Simulations	“I think part of it is it was at the top of the list. I went from the top down, and then I was trying to look underneath it instead of at the top of it.” “I think [finding the play button] would be easier if it was moved at the bottom of all the things that you select.”
	Post Processing	“I like the layout where it’s in the three-column design and the way it told me in the instructions to go to which column, that was helpful.”
	Visualizations	“It took me a second to figure out how to plot the map because [the instructions] said, ‘Make the map,’ and so I didn’t know. I was like, ‘Oh, what button do I click to make the map?’ I” read it like top of the map to the bottom. I, honestly, was lost. I was like, you know what, let’s just click this button and see what happens.” “I thought it was going to be two maps to subtract but I couldn’t get one on top of the other [in the map area]”, “‘top’

		and ‘bottom’ aren’t the best words to use. I would recommend, ‘First input’ because you do give numbers here. Maybe the numbers, ‘Graph 1,’ or ‘Graph 2’ I feel like that would be easier.”
B	Run Simulations	“I got confused because I was like, I know I saw [the proceed to post processing button].”
	Post Processing	“Fairly straight forward”
	Visualizations	“I think it helps that the data are numbered, 1. Annual surface with modern, 2. Annual surface with IPCC”.

Heat Maps and Gaze Plots

Heat maps were created using the aggregated fixations per page for all participants for each page (figure 9, figure 10, and figure 11). It is important to note that although they can be used to qualitatively assess patterns with the data, they do not encompass many of the nuances that are present in a dynamic website due to the fact that the fixations are placed over a stationary image of the website that likely does not represent the true configuration of the website throughout the task. This is because in a website that is interactive, elements on the screen are moving, and sometimes overlapping, which cannot be shown on the stationary placeholder image underlying the fixation data (Räihä, et al., 2005). Although the heat maps show slight differences between versions of the website on each page, the overall pattern between versions are fairly similar when looking at the heatmaps for each page. On the run simulations page (figure 9) for the (A) version, there is an area of heavy fixation at the bottom of the leftmost column where the record results button was located. On the (B) version, the record results button was removed, and the pause/play button was placed at the bottom of the column which also is an

area of heavy fixation. This suggests that users of this website will work down the column when completing the task from top to bottom regardless of if that is the correct area to run the simulation. In the (B) version, there is heavier fixation on the instructions, although the statistics do not suggest a significant difference in the average fixation duration in the run simulations instructions between versions. Aligning with the statistical findings, the heat maps show less fixation on the proceed to post processing button and heavier fixation on the simulations menu for the (A) version of the website (table 4, figure 9).

The heat maps that show aggregated user fixations overlaid on an image the post processing page, have some qualitative differences even though there were no statistical differences between groups (figure 8). The (B) version of the website appears to have heavier fixations over the instructions compared to group (A). Across versions, the pattern of fixations are similar for where users generally looked with most users looking heavily at the middle of the instructions and the areas necessary to complete the task, such as the selection menu. The heat maps for the visualizations page have very few qualitative differences (figure 10). It appears that in the (A) version there is slightly heavier fixation over the area of the instructions that asks users to select an operation than there is in the (B) version. Additionally, it appears that there is a slightly larger area of fixation on the (B) version near the operations menu and create map button than there is over the same area in the (A) version, however, it is difficult to discern if these fixations are on the mapping buttons or the operations menu due to the overlapping nature of these areas when they are in use and the limitation that a heat map is shown on a fixed area.



Figure 9: Heat maps showing aggregated fixations for all participants viewing the original version (A) and redesigned version (B) of EzGCM for the run simulations page.

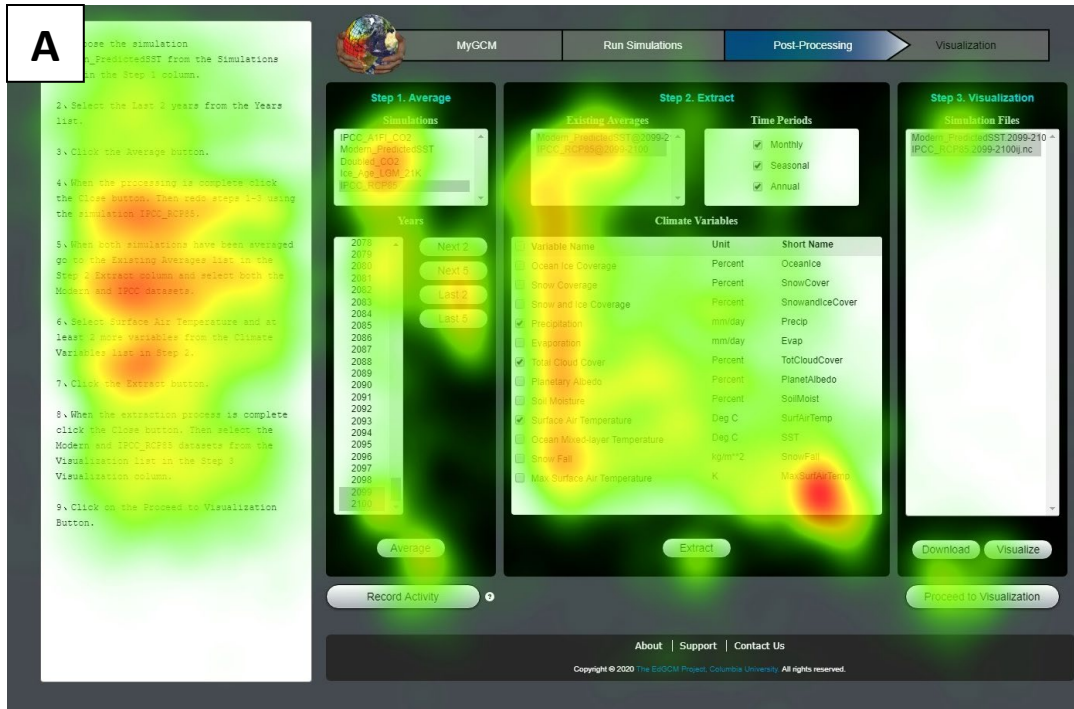


Figure 10: Heat maps showing aggregated fixations for all participants viewing the original version (A) and redesigned version (B) of EzGCM for the post processing page.

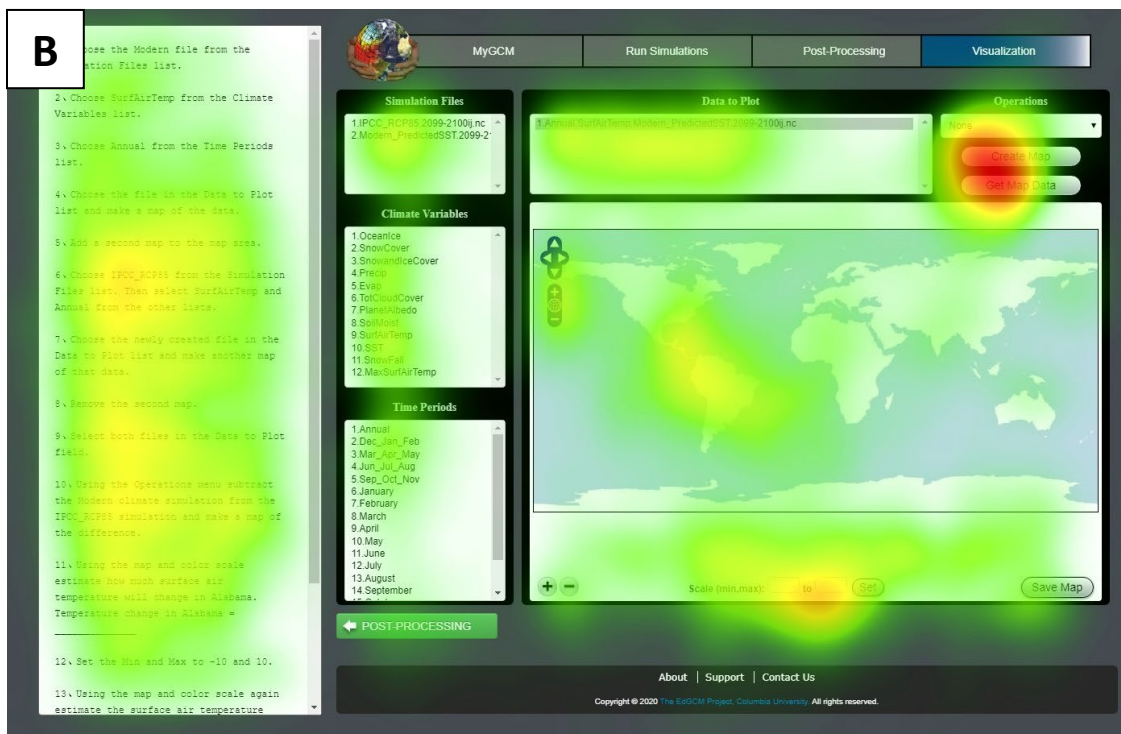
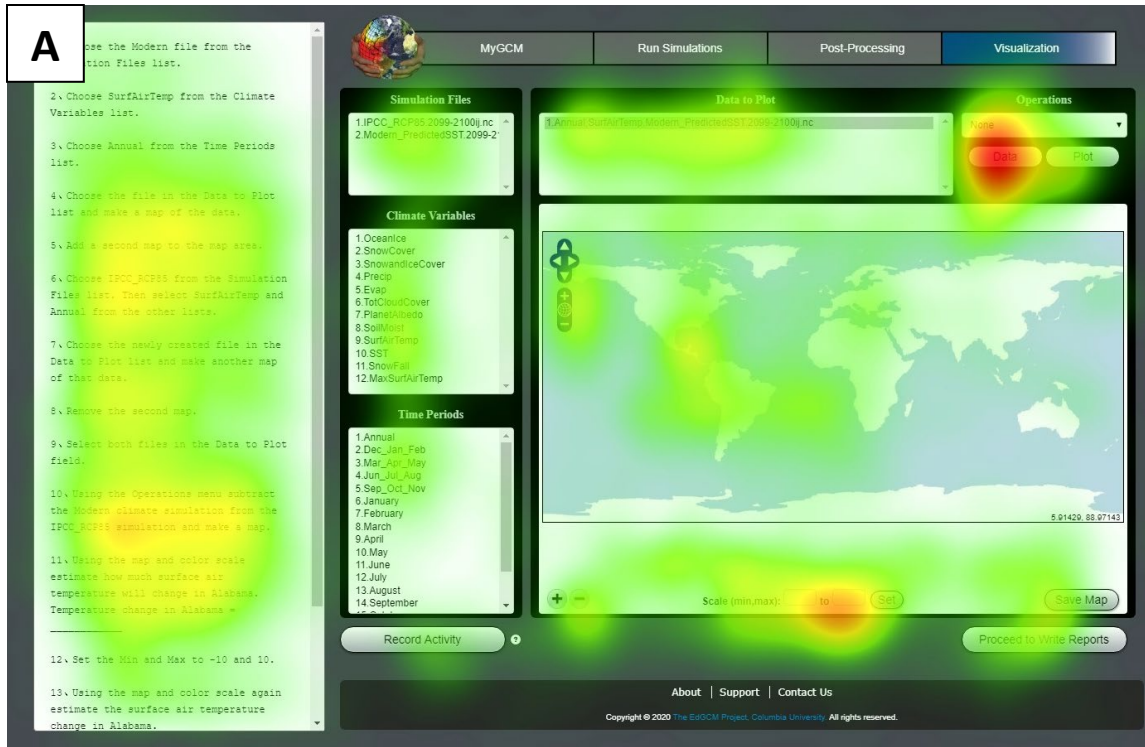


Figure 11: Heat map representing aggregated fixations for all participants viewing the original (A) and redesigned (B) version of EzGCM for the visualizations page.

Gaze plots were also created using the data from a single participant in each group (figure 12 and figure 13). Gaze plots for each page reveal that participants oscillated between the instructions and the page constantly. Also, for the run simulations page, it shows a downward scanpath for both versions where participants moved from the simulations page down to the bottom of that column no matter which version they had. The participant in the (A) version then went back to the instructions while looking for the pause and play button. In addition, on both versions, participants scan paths are going back and forth between the navigational bar and the proceed button, however, the participant in the (B) version had less fixations on the navigational bar. On the visualizations page, there were more fixations in the data to plot menu for the (A) version than the (B) version. Both participants had many scanpaths connecting the instructions, particularly where the operations related instruction is, and the plot/create map button.

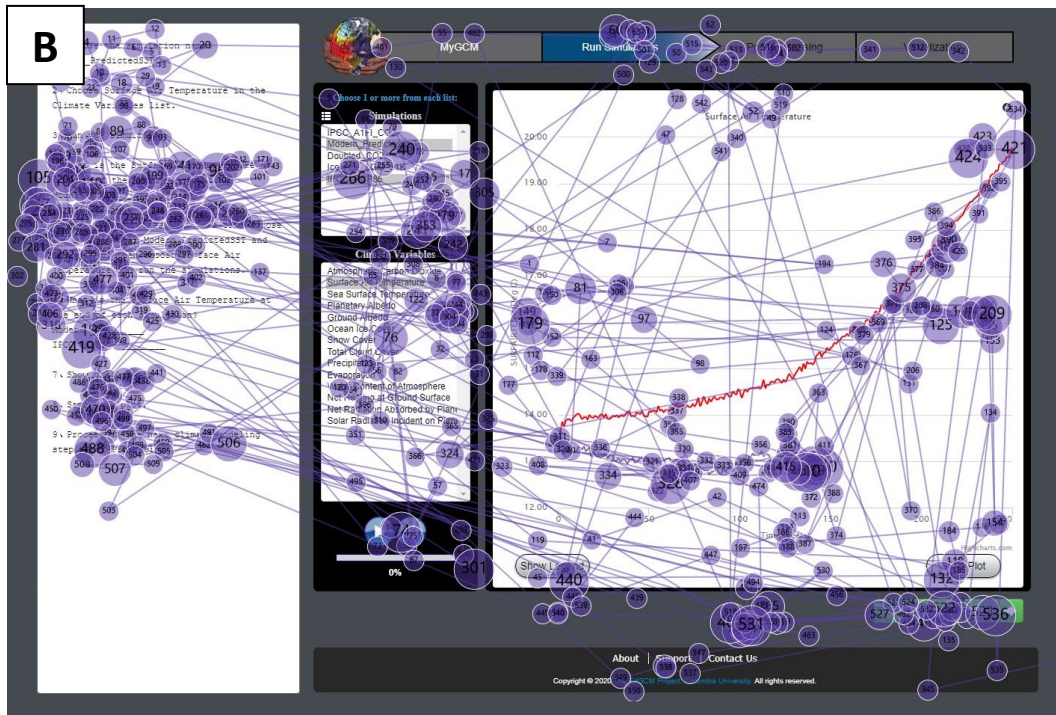
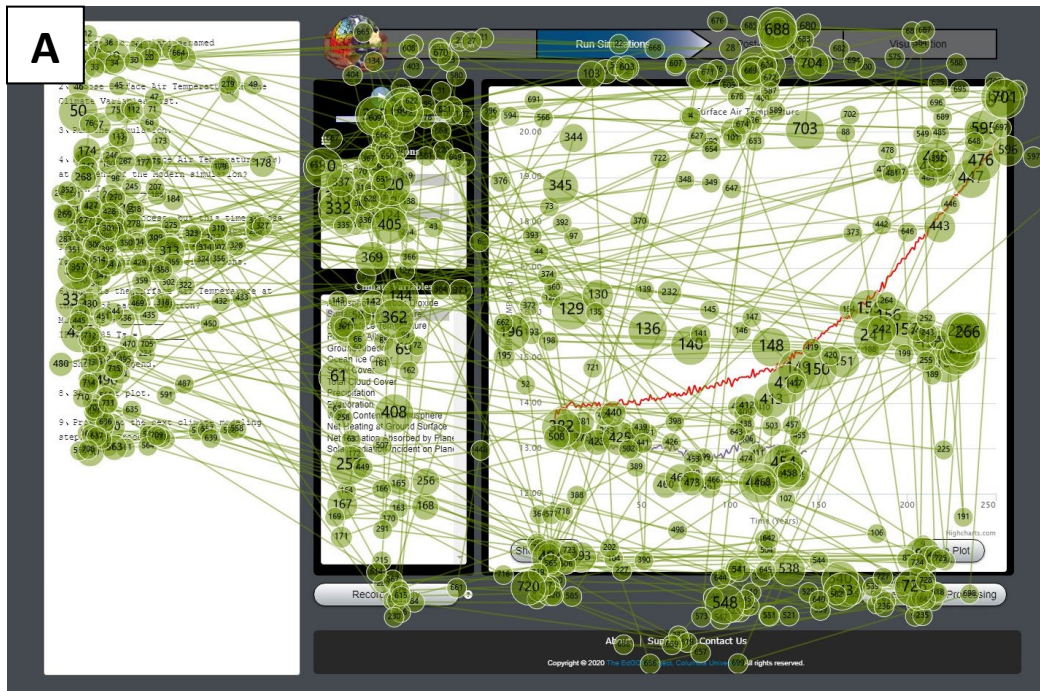


Figure 12: Gaze plots representing the scan path of one participant while navigating the run simulations page of the original version (A) and the redesigned version (B) of EzGCM.

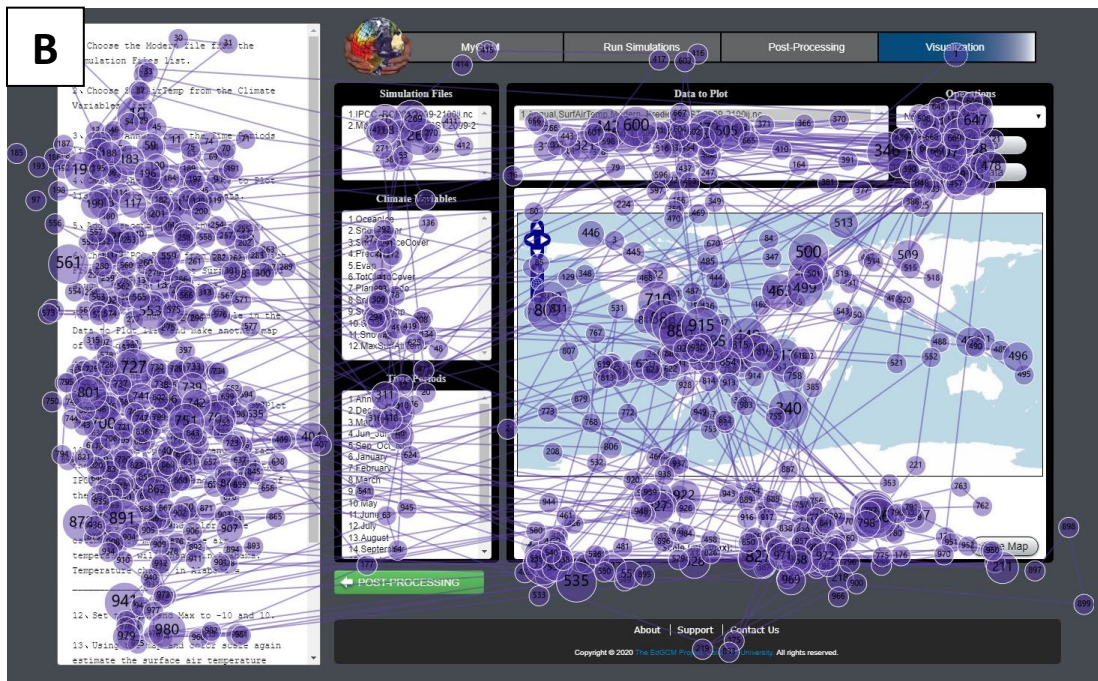


Figure 13: Gaze plots representing the scan path of one participant while navigating the post processing page of the original version (A) and the redesigned version (B) of EzGCM.

DISSCUSSION

This study adopted a three-pronged approach to evaluating the usability of two versions of an educational climate related webtool by investigating efficiency, accuracy, and satisfaction of users. This methodological approach has been established by other researchers in the field who have investigated the usability of other science related tools, including decision support tools, and science related website, much like the one used in the current study using these same three dimensions of usability (Cashwell, 2023; Maudlin, 2018). By looking at various data sources, researchers are able to establish a holistic view of the user experience of a tool and provide more useful information to stakeholders on how to improve the product. The following discussion will be framed around the evaluation of the efficiency, accuracy, and satisfaction for users of both versions of the tool following the relevant AOIs on each page of the website in order.

Run simulations page

Proceed to post processing button

Eye tracking metrics were used to compare the efficiency of the original and redesigned versions of EzGCM for AOIs drawn on each page. On the run simulations page, the analysis revealed that participants spent significantly less time on average looking at redesigned proceed to post processing button. Other studies with similar findings suggested that the average fixation duration can be an indication of the cognitive processing, with longer duration indicating more processing (Cooke, 2006; Just and Carpenter 1976). Another interpretation is that longer average fixation duration indicates that participants are struggling with information extraction (Fitts et al. 1950). When redesigning the button, researchers wanted to utilize visual cuing, which can

include the use of a number of design choices including color, arrows, variable font size, which can lessen the cognitive load of users (Richardson et al., 2009). In addition to visual cuing, the shorter average fixation duration on the green button could be a result of priming. The redesigned button, which advances users to the next page, was chosen to be green which has the perception of meaning “go” in society (Johnson, 2020). Because users already had an idea of what the button may do from the color of the button, paired with the visual cue of an arrow, they spent less time processing, and thus, less time fixating. Contrastingly, the original version, which utilized a grey button without symbology, required users to spend a longer amount of time extracting the information from that button before completing the task and advancing to the next page. Although there was a significant difference in total fixation duration between versions for this button, there was not a significant difference in the amount of time it took users to click the button between versions. This may be explained by the fact that as part of the task, users had to show the legend of the plot which hid the proceed to post processing button. On both versions of the webtool, this frustrated users as they were unsure how to proceed to the next page, even if they had seen the proceed button previously. When reflecting during the interview, one participant in the B group stated, “I got confused because I was like, I know I saw it.” While referring to the proceed button being hidden by the legend. Although the redesigned button proved to be more efficient than the original design based on average user fixation, future iterations of the website should avoid covering action buttons in order to further increase the efficiency and satisfaction of users.

Pause/Play Button

The other change made on the run simulations page was to the pause and play button, which was moved from the top of the column in the original version to the bottom of the column

in the new version. This change was made in support of prior eye tracking research which suggests that in general, people navigate pages from top to bottom in an F or triangular pattern. (Buscher et al, 2009). The hope was that moving the buttons below the operations will signal to the user that they need to select from the simulations menu and then will continue in a downward motion to use the buttons to run the simulation. To measure the efficiency of this change, we used time to first click and total fixation duration in the related AOIs. Eye tracking analysis revealed that there was no significant difference in the total fixation duration for the AOI containing the pause and play controls between versions. However, this may be explained by the fact that the button itself was not changed, just the location of the button. Once users in either group were able to locate the button, the amount of time processing the information from the button, or fixating on the button, should not significantly differ between groups.

Simulations Menu

Perhaps more explanatory than the time spent by users fixating on the button itself is the time they spent in the simulations menu, which is the step that must be completed immediately prior to clicking the pause and play button. Although the simulations menu did not change between versions, it was found that that users in group (A) spent significantly more time looking in the simulations menu than those who received the (B) version of the website. This could indicate that users who received the (A) version of the website selected from the simulations menu and then continued to fixate in that area while they looked for the pause play button to run the simulation which supports the findings of Ehmke and Wilson (2007) from their assessment of common usability problems in relation to eye tracking which argues that a longer fixation duration is indicative of confusion. Contrastingly, the reason that those who received the (B) version spent much less time fixating in the simulations menu may be that they were able to

move on to the next action of clicking the play button more efficiently. Many participants in the (A) group stated in their follow up interview that they were confused when finding the pause play button. When asked to explain their thought process while looking for the play button, one user stated, “I think part of it is it was at the top of the list. I went from the top down, and then I was trying to look underneath it instead of at the top of it.”. Another suggested, “I think [finding the play button] would be easier if it was moved at the bottom of all the things that you select.”. The qualitative data coupled with the quantitative eye tracking data suggests that the redesigned version of the run simulations page was more efficient and had a greater participant satisfaction than the original version.

Post Processing Page

Instructional Detail

On the post processing page, the level of detail within the instructions varied with the original version receiving more point and click instructions while the new version received more vague instructions. Despite these changes, there was not a significant difference in total fixation duration on the instructions panel present on the post processing page between groups. Similarly, there was no statistical difference in time to first click on the proceed to visualizations button, which serves as a measure for the total amount of time spent on the page since a click in that AOI advances users to the next task. This suggests that users were able to complete the task regardless of the detail in the instructions on this page. Although it may require further investigation on whether the level of instruction impacts user experience for this tool, the lack of statistical difference between groups on this page where minimal changes were made shows that there were also little to no differences in participant engagement between treatments. This finding supports that differences seen on other pages likely are due to changes made between

treatments. Many participants during the interview claimed that the page was “fairly straight forward” and often cited the column layout design as helping with navigation. One participant stated, “I like the layout where it's in the three-column design and the way it told me in the instructions to go to which column, that was helpful.”. This finding is aligned with other research in the field such as a study conducted by Namoun and colleagues (2018) which concluded that organizing content within column-based web layouts compared to a grid layout expedite successful target location and creates a positive user experience.

Proceed to Visualizations Page

Additionally, users interacted with the proceed to visualizations button in a similar fashion to the way they interacted with the proceed to post processing button on the run simulations. Although not significant at the .05 significance level, the differences in averages between groups suggested that users fixated longer on the original proceed button than they did on the redesigned version. This further supports that the use of visual cues, like arrows and colors, lessens the cognitive load of users for the same task.

Visualizations Page

Create Map/Plot Button

On the visualizations page, the average fixation duration on the button “create map” in the (B) version was significantly longer than the “plot” button in version (A). When designing an element of a website, research suggests that there is a careful balance between clarity and having too many words that will increase cognitive load (Bojoko, 2013). In the case of this website, the difference in fixation time on the button may be explained by the increased amount of time that it took users to read the label on the redesigned version. In addition, the new button was much

larger in size than the old button. In general, larger elements will capture more of a user's attention than smaller elements which may explain the longer fixation time on the (B) version.

Although fixation duration was longer, the time to first click results between groups revealed that users were able to accomplish the first task of making the single variable map faster when they received the new version of the website. Supporting this statistical evidence, one user in group A stated in their interview "It took me a second to figure out how to plot the map because [the instructions] said, 'Make the map,' and so I didn't know. I was like, 'Oh, what button do I click to make the map?'" These findings suggest that the word "plot" is not as clear to users as "create map" in relaying the purpose of the button. In their study of common usability problems, Ehmke and Wilson (2007) describe the issue of unclear terminology, meaning that the wording or layout seems to be unclear as the requested information is displayed but the participant still goes further to another page. They describe that when this issue occurs, it may be characterized by fixation on target but further fixations across other elements on page and further scanning with regressions back to unclear element. If this were occurring in this study, this may explain why the time taken to click the "plot" button was longer than the "create map" because users were looking around at other elements before clicking the button due to being unfamiliar with the word.

Operations Menu

The changes made to the operations menu also seemed to affect the amount of time users spent before clicking in the menu. Participants who received the (B) version of the website spent less time before selecting an option in the menu. The increased amount of time on the original version may be the result of confusion with the terminology "top" and "bottom" within the menu which caused users to spend more time trying to understand before making a selection. One user

who received the original version of the website stated “I read it like top of the map to the bottom. I, honestly, was lost. I was like, you know what, let's just click this button and see what happens.” When asked about their rationale for selecting the operation they chose. Along the same lines, another participant added “I thought it was going to be two maps to subtract but I couldn't get one on top of the other [in the map area]”, which signals confusion as to the meaning of the menu. Another user even suggested, “‘top’ and ‘bottom’ aren't the best words to use. I would recommend, ‘First input’ because you do give numbers here. Maybe the numbers, ‘Graph 1,’ or ‘Graph 2’ I feel like that would be easier” when asked how they would improve the menu. Contrastingly, a user who received the (B) version of the website stated, “I think it helps that the data are numbered, 1. Annual surface with modern, 2. Annual surface with IPCC”.

Supporting the claim that users in group (A) were confused by the operations menu is the data collected about the accuracy on the visualizations page. Participants were asked to create a map and answer a related question, which required them to successfully use the operations menu. The accuracy for the question increased by 36% when comparing the original version of the website to the new version of the website. This difference in correctness was significant at the .05 confidence level. Similarly, the accuracy of the saved maps differed by ~22% between groups. The intersection of accuracy, efficiency, and satisfaction all suggests that the redesigned operations menu outperformed the original. It is recommended that the verbiage in the operations menu be changed if this tool were to be used without formal instruction that explained the meaning of that terminology within the menu.

Overall satisfaction

In addition to the qualitative data from the interviews regarding satisfaction, the analysis of post surveys yielded some important information about between version comparisons. Neither

group was likely to recommend this website, with that question receiving the lowest average scores. This is likely due to the website not being applicable to most of the participants in the study on a practical basis. Overall, users answered more positively on the satisfaction post survey if they received the redesigned version. Users in the (B) version group felt more strongly that this website could be used by people with no prior climate data which may be indicative of a lessened level of confusion and frustration since the participants in this study were generally unfamiliar with climate data. It is also important to note that the overall ranking for ease of navigation was greater when users were given the (B) version of the website. In addition to satisfaction, users who received the (B) version were statistically more likely to say that their perception of climate science differed from the pre survey. This may be explained by the increased overall accuracy on the website by users in this group, increased efficiency with less user confusion, and general satisfaction with this version of the website. Because the main goal of teaching climate modeling with this website is to encourage students to engage in the scientific process and demystify the projections created by climate scientists, an overall increased perception is an incredibly important finding in this study. This supports that websites with greater usability are able to accomplish their goals and impact a user more adequately.

CONCLUSIONS AND RECOMMENDATIONS

The primary goal of usability testing is to identify usability issues and to provide recommendations on how to improve the product (Nilsen and Landauer, 1993). Analyzing the three dimensions of usability is important because although they may yield similar conclusions, one metric may not be enough to gain a sense of user experience if there is little or no difference between the versions. For example, accuracy for the question and saved product for the run

simulations did not differ between versions. However, it is clear from the eye tracking and interview data that both efficiency and satisfaction were greater when users received the redesigned version of the website.

Based on the analysis performed in this study, researchers recommend that future iterations of this website should incorporate more visual cues, such as symbols and colors, into the web design in order to further increase user efficiency. In addition, other educational websites may also benefit from incorporating similar strategies into their design as well including the use of intentional color and symbology that may lessen the cognitive load of users. The results in this study also suggest the importance of using clear labels and avoiding terminology that may be unfamiliar to the website audience when creating elements of a website in order to increase efficiency. However, website designers also need to acknowledge the tradeoff between adding text to increase clarity which may decrease efficiency by increasing the cognitive load of participants.

Based off of the usability testing in this study, it is clear that the design changes made to version (B) increased efficiency in regard to fixation duration and time to first click. Although these changes may seem small, they drastically impacted participants' ability to correctly perform the task being asked. Additionally, the perceived satisfaction of users was greater overall when given the redesigned version of the website. For these reasons, researchers recommend that the stakeholder in this study adopt the changes made to EzGCM in the redesigned version and consider adding more visual cues and clearer labels in the future.

LIMITATIONS & FUTURE DIRECTIONS

Although the current study was able to make suggestions about how to improve the usability of EzGCM, there were some limitations that deserve to be addressed and questions to investigate in future studies. Although the original goal was to recruit 30 per version group, eye-tracking studies typically tend to have smaller sample sizes, so the amount collected was considered a reasonable number of participants to draw conclusions about the usability of this website (Jacob & Karn, 2003; Alaçam & Dalc, 2009; Bucher, & Schumacher, 2006). In addition, there are many limitations inherent to visualizing eye tracking data that is collected from a dynamic stimulus. The two most common methods for visualizing eye tracking data, the heat map and the gaze plot, both involve placing collected data on a static image (Bojko, 2013). For heat maps of a dynamic website, like the one in this study, this neglects the spatial distribution of fixations when there are overlapping areas of interest or data that changes temporally. Gaze plots have the same limitations plus for data collected for longer tasks, they are often cluttered and indiscernible when it comes to trends due to the high number of overlapping fixations. There are some researchers who have presented alternative ways to view scanpath data, like Rähä and colleagues who presented the idea of the time plot, however, these visualizations are time consuming to construct and still present issues for dynamic stimuli (Rähä, et. al., 2005).

Future studies should explore innovative data visualizations for the data collected from interactive websites which may uncover new findings about trends in how users navigated this website. Although this study explored general patterns and made suggestions about overall usability, more conclusions may be made by looking at the eye tracking data from subsets of the population sample. For example, future studies should investigate differences between users who completed the task accurately and those who failed to complete the task accurately, STEM vs.

non-STEM majors, or even males vs. females to understand if usability of this website differs among various groups. In addition, future studies should further investigate the impact of prior climate knowledge and/or various demographic characteristics of participants collected on the pre survey may have on user's abilities to navigate the tool or their satisfaction with the tool through regression analysis. Although significant results were highly discussed in this study, it may also be of use to investigate AOIs that were not heavily fixated in, such as the add map button, and redesign them in future iterations of the webtool. Finally, further iterations and designs for the website should also undergo usability testing in order to ensure the most user-friendly product for undergraduates in the classroom.

SIGNIFICANCE:

Undoubtedly, climate change will bring socioeconomic and environmental challenges and opportunities; citizens who understand climate science will be better prepared to respond to both (IPCC, 2014). Science tools, like EzGCM, have the ability to improve the climate literacy of students and thus make them more informed consumers of climate information to intelligently engage in public discourse surrounding the issue (Van Nuland, Hall, & Langley, 2020; Develaki, (2017). Although climate education is incredibly important, there are few resources available to educators available to teach about it. When webtools are user friendly, it encourages adoption into the curriculum and positively impacts student outcomes from the tool. Broadly, lessons learned about the features of EzGCM that positively impacted the three dimensions of usability for users of this website, such as button placement, button design and language choices, will help to fortify user-centered design frameworks that can be incorporated into new scientific pedagogical tools.

REFERENCES

- Alaçam, Ö., & Dalcı, M. (2009, July). A usability study of webmaps with eye tracking tool: The effects of iconic representation of information. In *International conference on human-computer interaction* (pp. 12-21). Springer, Berlin, Heidelberg.
- Bhattacharya, D., Carroll-Steward, K., Chandler, M., & Forbes, C. (2020). Investigating the phenomenon of increasing surface air temperatures using a global climate modeling approach.
- Bhattacharya, D., Steward K. C., & Forbes, C. T. (2021) Climate education in secondary science: comparison of model-based and non-model-based investigations of Earth's climate, *International Journal of Science Education*.
- Brown, N. J., Furtak, E. M., Timms, M., Nagashima, S. O., & Wilson, M. (2010). The evidence-based reasoning framework: Assessing scientific reasoning. *Educational Assessment*, 15(3-4): 123-141.
- Bojko, A. (2013). *Eye Tracking the User Experience: A Practical Guide to Research*. New York, NY: Rosenfeld Media.
- Bucher, H. & Schumacher, P. (2006). The relevance of attention for selecting news content. An eye-tracking study on attention patterns in the reception of print and online media., 31(3), 347-368. <https://doi.org/10.1515/COMMUN.2006.022>
- Buscher, G., Cutrell, E., and Morris, M.R. (2009) What do you see when you're surfing? Using eye tracking to predict salient regions of web pages. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, 4-9*. pp. 21-30.
- Cashwell, Haven J., et al. "User Engagement Testing with a Pilot Decision Support Tool Aimed to Support Species Managers." *Weather, Climate, and Society* 15.2 (2023): 327-338.
- Connor, C.E., Egeth, H.E., and Yantis, S. (2004) Visual attention: Bottom-up versus top-down. *Current Biology*. 14, R850-R852. DOI: 10.1016/j.cub.2004.09.041
- Cooke, L. (2006). Is Eye Tracking the Next Step in Usability Testing?. IEEE International Professional Communication Conference, International Professional Communication Conference, 2006 IEEE, 236. doi:10.1109/IPCC.2006.320355
- Cox, J. and Dale, B.G. (2002), "Key quality factors in Web site design and use: an examination", *International Journal of Quality & Reliability Management*, Vol. 19 No. 7, pp. 862-888. <https://doi.org/10.1108/02656710210434784>
- Develaki, M. (2017). Using Computer Simulations for Promoting Model-based Reasoning. *Science and Education* 26, 1001-1027
- Ehmke, C., & Wilson, S. (2007). Identifying web usability problems from eyetracking data. DOI: 10.14236/ewic/HCI2007.12

- Fitts, P. M., Jones, R. E., and Milton, J. L. (1950). Eye movements of aircraft pilots during instrument landing approaches. *Aeronautical Engineering Review* Vol. 9(2), pp.24–29
- Goldberg, J. H., and Kotval, X. P. (1998). Eye movement-based evaluation of the
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Halszka, J., and van de Weijer, J. (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press.
- Hung, J.C.; Wang, C.C. (2020). Exploring the website object layout of responsive web design: Results of eye tracking evaluations. *The Journal of Supercomputing*, 77, 343–365.
- IPCC (2014). *Climate Change 2014: Synthesis Report*. Geneva, Switzerland: IPCC.
- Johnson, J. (2020). *Designing with the mind in mind: simple guide to understanding user interface design guidelines*. Morgan Kaufmann.
- Just, M. A., Carpenter, P. A., & Carnegie-Mellon Univ., P. P. (1975). *Eye Fixations and Cognitive Processes*.
- Kim, M. C., Hannafin, M. J., & Bryan, L. A. (2007). Technology-enhanced inquiry tools in science education: An emerging pedagogical framework for classroom practice. *Science education*, 91(6), 1010-1030.
- McNeal, K., Libarkin, J. C., Tamara S. L., Bardar, E., Haddad, N., Ellins K., & Dutta S. (2014) The Role of Research in Online Curriculum Development: The Case of EarthLabs Climate Change and Earth System Modules, *Journal of Geoscience Education*, 62:4, 560-577
- Maudlin, Lindsay C., et al. "Website usability differences between males and females: An eye-tracking evaluation of a climate decision support system." *Weather, climate, and society* 12.1 (2020): 183-192.
- Maudlin, Lindsay Catherine. "An Evaluation of a Climate Decision Support System: An Eye-Tracking Study." (2018).
- Meadow, A. M., Ferguson, D. B., Guido, Z., Horangic, A., Owen, G., & Wall, T. (2015). Moving toward the deliberate coproduction of climate science knowledge. *Weather, Climate, and Society*, 7(2), 179-191.
- Namoun, A. (2018). Three Column Website Layout vs. Grid Website Layout: An Eye Tracking Study. In: Marcus, A., Wang, W. (eds) *Design, User Experience, and Usability: Designing Interactions*. DUXU 2018. Lecture Notes in Computer Science(), vol 10919. Springer, Cham. https://doi.org/10.1007/978-3-319-91803-7_20
- Nielsen, J, and Landauer, T. K. (1993). A mathematical model of the finding of usability problems. *Conference On Human Factors In Computing Systems Proceedings*, 206. doi:10.1145/169059.169166
- Pinto, Y., van der Leij, A., Sligte, I.G., Lamme, V.A.F., and Scholte, H.S. (2013). Bottom-up and top-down attention are independent. *Journal of Vision*, 13, 1–14.
- Pretorius, M. C., Calitz, A. P., & van Greunen, D. (2005, July). The added value of eye tracking in the usability evaluation of a network management tool. In *Proceedings of the 2005*

- annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries* (pp. 1-10).
- Räihä, K.J., Aula, A., Majaranta, P., Rantala, H., Koivunen, K. (2005). Static Visualization of Temporal Eye-Tracking Data. In: Costabile, M.F., Paternò, F. (eds) *Human-Computer Interaction - INTERACT 2005*. INTERACT 2005. Lecture Notes in Computer Science, vol 3585. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11555261_76
- Renshaw, J.A., Finlay, J.E., Tyfa, D., and Ward, R.D. (2003). Designing for Visual Influence: an Eye Tracking Study of the Usability of Graphical Management Information. In *proceedings of the IFP conference on Human-Computer interaction* (INTRACT 2003). IOS press, London. pp 144-151
- Richardson, R. T., Drexler, T. L., & Delparte, D. M. (2014). Color and contrast in E-Learning design: A review of the literature and recommendations for instructional designers and web developers. *MERLOT Journal of Online Learning and Teaching*, 10(4), 657-670.
- Shackley, S., Young, P., Parkinson, S., & Wynne, B. (1998). Uncertainty, complexity and concepts of good science in climate change modelling: are GCMs the best tools?. *Climatic change*, 38(2), 159-205.
- Spool, J., Scanlan, T., Schroeder, W., Snyder, C. and De Angels, T. (1999), *Web Site Usability: A Designers Guide*, Morgan Kaufmann Publishers, New York, NY.
- Steven M. Manson , Len Kne , Kevin R. Dyke , Jerry Shannon & Sami Eria (2012) Using Eye-tracking and Mouse Metrics to Test Usability of Web Mapping Navigation, *Cartography and Geographic Information Science*, 39:1, 48-60, DOI: 10.1559/1523040639148
- Țichindelean, M., Țichindelean, M. T., Cetină, I., & Orzan, G. (2021). A Comparative Eye Tracking Study of Usability—Towards Sustainable Web Design. *Sustainability*, 13(18), 10415.
- Van Nuland, S. E., Hall, E., & Langley, N. R. (2020). STEM crisis teaching: curriculum design with e-learning tools. *FASEB BioAdvances*, 2(11), 631-637.
- Vassilopoulou, K. and Keeling, K. (2000), “Usability measurement tool survey: differences between Australia and UK”, Manchester School of Management, UMIST, Manchester.
- Wang, J., Antonenko, P., Celepkolu, M., Jimenez, Y., Fieldman, E., & Fieldman, A. (2019). Exploring relationships between eye tracking and traditional usability testing data. *International Journal of Human–Computer Interaction*, 35(6), 483-494.
- Yuen, S. C. Y., Yaoyuneyong, G., & Yuen, P. K. (2011). Perceptions, interest, and use: Teachers and web 2.0 tools in education. *International Journal of Technology in Teaching & Learning*, 7(2).
- Zardari, B. A., Hussain, Z., Arain, A. A., Rizvi, W. H., & Vighio, M. S. (2021). QUEST e-learning portal: Applying heuristic evaluation, usability testing and eye tracking. *Universal Access in the Information Society*, 20(3), 531-543.

APPENDIX A

Table 1: Timeline of usability study

Task	Tentative Date
Ongoing meetings with stakeholder to decide on website changes	Fall 2021-Summer 2022
Thesis Proposal	Spring 2023
Software developers making beta version	Spring 2022-Fall 2022
Participant recruitment and eye-tracking study	Spring 2023
Data analysis and writing	Spring-Summer 2023
Defense	Summer 2023

Table 2: Tasks for each page

Page	Version	Task
Run Simulations	(A) and (B)	<ol style="list-style-type: none"> 1、 Choose the simulation named Modern_PredictedSST 2、 Choose Surface Air Temperature in the Climate Variables list. 3、 Run the simulation. 4、 What is the Surface Air Temperature (Ts) at the end of the Modern simulation? Modern Ts = _____ 5、 Repeat the process, but this time choose two simulations, Modern_PredictedSST and IPCC_RCP85, then choose Surface Air Temperature and run the simulations. 6、 What is the Surface Air Temperature at the end of each simulation? Modern Ts = _____ IPCC_RCP85 Ts = _____ 7、 Show the legend. 8、 Save your plot. 9、 Proceed to the next climate modeling step, Post-Processing.
Post processing	(A)	<ol style="list-style-type: none"> 1、 Choose the simulation Modern_PredictedSST from the Simulations list in the Step 1 column. 2、 Select the Last 2 years from the Years list. 3、 Click the Average button. 4、 When the processing is complete click the Close button. Then redo steps 1-3 using the simulation IPCC_RCP85.

	<p>5、 When both simulations have been averaged go to the Existing Averages list in the Step 2 Extract column and select both the Modern and IPCC datasets.</p> <p>6、 Select Surface Air Temperature and at least 2 more variables from the Climate Variables list in Step 2.</p> <p>7、 Click the Extract button.</p> <p>8、 When the extraction process is complete click the Close button. Then select the Modern and IPCC_RCP85 datasets from the Visualization list in the Step 3 Visualization column.</p> <p>9、 Click on the Proceed to Visualization Button.</p>
(B)	<p>1、 Choose the simulation Modern_PredictedSST from the Simulations list in the Step 1 column.</p> <p>2、 Select years 2099-2100 from the Years list.</p> <p>3、 Average the last 2 years.</p> <p>4、 When the averaging is complete close the window and redo steps 1-3 using the simulation IPCC_RCP85.</p> <p>5、 When both simulations have been averaged select both the Modern and IPCC files from the Existing Averages list in the Extract column.</p> <p>6、 Select Surface Air Temperature and at least 2 more climate variables.</p> <p>7、 Extract the data.</p> <p>8、 When extracting is complete close the window and select the Modern and IPCC_RCP85 files from the Visualization list.</p> <p>9、 Proceed to Visualization.</p>

Visualizations	(A) and (B)	<ol style="list-style-type: none"> 1、 Choose the Modern file from the Simulation Files list. 2、 Choose SurfAirTemp from the Climate Variables list. 3、 Choose Annual from the Time Periods list. 4、 Choose the file in the Data to Plot list and make a map of the data. 5、 Add a second map to the map area. 6、 Choose IPCC_RCP85 from the Simulation Files list. Then select SurfAirTemp and Annual from the other lists. 7、 Choose the newly created file in the Data to Plot list and make another map of that data. 8、 Remove the second map. 9、 Select both files in the Data to Plot field. 10、 Using the Operations menu subtract the Modern climate simulation from the IPCC_RCP85 simulation and make a map. 11、 Using the map and color scale estimate how much surface air temperature will change in Alabama. Temperature change in Alabama = _____ 12、 Set the Min and Max to -10 and 10. 13、 Using the map and color scale again estimate the surface air temperature change in Alabama. Temperature change in Alabama = _____ 14、 Save your map. 15、 Press the F10 key on your keyboard.
----------------	-------------	---

Table 3: Statistical output for between group t tests.

Metric	AOI	Avg OG	Std dev	Avg NEW	Std dev b	Avg Diff	t	df	pvalue
Total	Add Map	1.51	2.14	1.31	1.22	0.21	0.33	26.21	0.7413
Fixation Duration	Instructions	70.77	31.68	69.75	25.04	1.03	0.12	41.19	0.905
	PP								
	Instructions	63.70	26.35	55.76	17.06	7.94	1.22	39.75	0.2286
	RS								
	Instructions	103.22	51.44	97.87	28.25	5.35	0.44	36.32	0.6611
	Vis								
	Map area	49.57		57.77		-8.20	- 0.81	42.32	0.4252
	Operations	7.51	5.70	5.99	4.45	1.52	0.98	39.48	0.3355
	Pause/Play	2.43	2.22	3.13	2.34	-0.71	- 0.96	35.44	0.3416
	Plot/Create	1.50	1.12	4.88	4.89	-3.38	- 3.04	23.25	0.005756
	Post Process	122.71	29.44	127.91	26.36	-5.21	- 0.62	41.99	0.5391
Proceed to PP	3.18	2.49	1.34	0.99	1.84	3.24	29.86	0.002922	
Proceed to vis	1.24	0.70	0.81	0.68	0.44	1.87	32.27	0.0704	

	Sims	43.50	14.61	32.99	9.95	10.51	2.87	40.74	0.006447
Time to First Click	Add Map	828.04	225.10	804.88	155.25	23.16	0.31	22.14	0.7576
	Map area	791.57		795.32		-3.75	- 0.08	45.01	0.9353
	Operations	207.89	116.42	94.66	207.11	113.23	2.11	27.41	0.04391
	Pause/Play	202.51	59.66	195.52	79.56	6.99	0.35	40.05	0.7294
	Plot/Create	96.18	100.39	46.48	30.65	49.70	2.17	23.87	0.04061
	Post Process	440.06	92.89	422.81	84.79	17.25	0.68	46.35	0.5008
	Proceed to PP	390.20		404.79		-14.59	- 0.85	42.94	0.3991
	Proceed to vis	675.97		653.68		22.29	0.67	40.95	0.5057
	Sims	160.22		178.65		-18.43	- 0.98	29.00	0.3335

APPENDIX B

Script for Recruitment in Undergraduate Classrooms:

Hello everyone. My name is Ally Brown and I am a graduate student obtaining my master's degree in geography in the Department of Geosciences. I am currently working with my advisor, Dr. Karen S. McNeal, on some research and we need your help in our research efforts. This research involves eye-tracking a NASA developed web-based climate modeling tool. We want to figure out the effectiveness and usability of this tool, so that we can help the designers make design changes to the site that will make it more effectively used by educators to teach students about the mechanics of climate modeling.

If you decide to participate in the research, you will have tasks to complete during the eye-tracking, and a subsequent interview. The research will be conducted in the Science Center Classroom Building and will take about one hour of your time. You will receive a \$25 gift card if you participate in the study. If you want to participate, please fill out this form ***insert link here*** to provide your name, email, and availability and Dr. Karen S. McNeal or myself will be in contact soon about setting up a date and time for the study.

If you have any questions about the study, you may ask them now or email me at JZB0184@auburn.edu or Dr. Karen S. McNeal at ksm0041@auburn.edu.

Thank you all so much for your time.

Recruitment Email for Undergraduates:

Hello,

The Auburn Geocognition Lab is conducting research and we need your help! We are studying a NASA developed online climate modeling tool that is meant to help students like you learn the mechanics behind climate models. This will be an eye-tracking study which will require 60-75 minutes of your time. This time will include the eye-tracking and an interview. If you choose to participate, you will be compensated for your time with a \$25 gift card.

If you are an undergraduate student, who is at least 19 years old, and would like to participate, please fill out the following form ***insert link here*** to provide your name, email aswell as your availability to participate in the in-person research. If you fill out the form, expect an email from me with details about the date and time for your in-person research.

Please contact Jena (Ally) Brown at JZB0184@auburn.edu or Dr. Karen S. McNeal at ksm0041@auburn.edu if you have any further questions.

Interview script:

“Hello, my name is Ally and I am going to walk you through today’s session. As you read in the informed consent document, I am a graduate student and NSF national research trainee currently working the developers of an online climate modeling tool to help improve the usability of this tool for educators and students in the classroom and ultimately inform a redesign of the website.

I’d like to begin by thanking you for making time to speak with me this afternoon. Your feedback is incredibly valuable and will help us determine if the website functions as intended. I just want to emphasize that there are no wrong answers during this interview and your honest feedback, negative and/or positive, is crucial to help make this website the most useful tool it can be for educators and students. Just to confirm, we would like to keep this session, both the eye tracking and interview to around 60-75 minutes. Does that still work for you?

Great. If you need a break or wish to stop at any time, please let me know.

During this session, I will start by asking you a few questions around you and your background. Later, I will ask you to sit at the desktop which will calibrate your eyes and give me some data afterwards about where, when, and for how long you are looking at the website. After accomplishing a few tasks using the website we are evaluating, I will ask you some more in detail questions about your experience and impressions of the website.

If you have any questions as we go along, just ask. I may not be able to answer them right away since we are interested in how people do when they don’t have someone sitting next to them to help. But if you still have any questions when we are done, I’ll try to answer them then.

With your permission, I would like to record this interview. The recording will be transcribed, then destroyed and not attached to any identifiable information.

At this time, I would like to invite you to ask any questions you have about my research, eye tracking, or the interview process. Does everything outlined in the document and reiterated here sound agreeable to you?

If any follow up questions arise surrounding this interview, please do not hesitate to contact me or my professor using the contact information listed on the form I just gave you.”

APPENDIX C

Pre-Survey Questions:

1. Age: _____
2. Year: _____
3. Gender:
 - a. Male
 - b. Female
 - c. Prefer not to answer
4. Race/Ethnicity:
 - a. American Indian/Native American/Alaskan Native
 - b. Asian/Asian American
 - c. Black/African/African American
 - d. Latino(a)/Hispanic
 - e. Native Hawaiian/Pacific Islander
 - f. White/Caucasian
 - g. Other:_____
5. Major: _____
6. How many classes have you taken about climate change or climate modeling?
 - a. 0
 - b. 1
 - c. 2
 - d. 3
 - e. 4

f. 5

g. > 5

7. How would you rank your confidence navigating web tools on the internet?

a. Not very confident

b. Fairly confident

c. Neither confident of unconfident

d. Confident

e. Very confident

8. How often do you use climate data?

a. Once a year

b. Once a month

c. Once a week

d. Once a day

e. More than once a day

f. I do not use climate data

9. Do you see climate change as a priority?

a. Low priority

b. Somewhat priority

c. Neutral

d. Moderate Priority

e. High Priority

Post-Survey Questions:

1. How likely are you to recommend EzGCM to someone else?
 - a. Extremely unlikely
 - b. Unlikely
 - c. Neutral
 - d. Likely
 - e. Extremely likely
2. Your knowledge about climate change improved after using EzGCM.
 - a. Disagree
 - b. Somewhat disagree
 - c. Neither agree or disagree
 - d. Somewhat agree
 - e. Agree
3. You feel like EzGCM can be used by people with no prior experience with climate data.
 - a. Disagree
 - b. Somewhat disagree
 - c. Neither agree or disagree
 - d. Somewhat agree
 - e. Agree
4. You feel as though EzGCM was easy to navigate.
 - a. Disagree
 - b. Somewhat disagree
 - c. Neither agree or disagree

d. Somewhat agree

e. Agree

5. Do you feel like your opinion on the priority of climate change differs from the pre-survey?

a. Yes

b. No

Interview Sample Questions:

1. If you were to rank your overall satisfaction with EzGCM from 1-10 (10 highest), what would you give it and why?
2. Can you identify any features that added or detracted from your experience navigating these tasks?
3. What did you think was the most difficult aspect to understand on EzGCM?
4. What did you think was the easiest aspect to understand on EzGCM?
5. Do you recommend any improvements for this framework?
6. If you were to change anything about EzGCM, what would you change?