Using Eye-Tracking to Test and Improve Website Design

Anna Prisacari and Thomas Holme

Human-Computer Interaction Program, Iowa State University, Ames, IA 50011, USA
{annacari,taholme}@iastate.edu

Abstract. In developing a website, it is essential to test its design. For example, users may look at a certain image or text paragraph without paying attention to what designers may consider being the most essential information or the users may erroneously interpret its design and get confused. If users don't interact with the website as designers anticipate, the design of website becomes dubious. In our eye-tracking study we invited 11 undergraduate students from an introductory chemistry course to test the usability of newly developed website on climate change. The results show that animated features draw more attention regardless of strength of relationship to content. Based on quantitative and qualitative data, we present possible recommendations how to improve the design of the website and how to enhance user's overall experience.

Keywords: eye-tracking, usability, web site design.

1 Introduction

One of the more challenging characteristics associated with the development of educational web sites is the wide array of experience and proclivity of those who carry out the work. Particularly for college-level target audiences, the people who envision and oversee development are often more likely to be content experts than they are design experts. Thus, educational materials development for the web has a wide range of probable effects on target audiences, and estimating the actual impact of these materials provides an important avenue for usability. Academic development efforts seldom include funding for usability testing while sites are being built, so empirical studies that investigate active sites developed in academic settings are an important mechanism for determining how to promote good design characteristics in web site development.

While this premise could be applied within any academic discipline, an overall interest in educational efforts within Science, Technology, Engineering and Mathematics (STEM) suggests that efforts to consider usability of educational materials for the web in these areas are particularly important. An emerging emphasis on interdisciplinary efforts within STEM education, in particular, provides additional motivation to consider web sites related to topics such as Climate Change, which are not only important in a global policy sense, but challenging because of the fundamental interdisciplinary nature of the topic. The study reported here considers the usability of an academically produced site about the chemistry aspects of climate change as an example of the type of empirical data that can be considered within this topic from the perspective of human-computer interactions.

A. Marcus (Ed.): DUXU/HCII 2013, Part I, LNCS 8012, pp. 389–398, 2013.

2 Literature on Usability and Eye-Tracking

Studies in human-computer interaction that probe usability of web site and factors of design that affect usability have become increasingly common. These studies have become more important in the academic environment as the number of students in higher education with reasonable information technology (IT) access continues to grow [1]. One key area of concern for academic uses of IT and internet information resources lies in how well students can locate needed information, a type of task that is strongly associated with the concept of information architecture [2,3]. Within this area, the concept of the depth of information processing undertaken, in particular, plays a prominent role in the manner in which students in academic environments use web sites [4]. Specific studies of the role of navigation elements compliment the concept of depth of information and information architecture conceptualizations of the challenge of presenting high information content density in web environments [5]. More recently, the interplay between practical usability and aesthetic design concepts have been noted as well [6,7], studies that may inform some aspects of the presentation of visually oriented information content in the sciences, for example.

Sites designed for students to engage in STEM materials are often infused with visual elements that scientific experts use to organize and summarize large amounts of information. Studies that focus on this type of visual representation are increasingly considering the challenges students face in digesting the information in these environments [8,9], particularly when the learning is self-directed or only lightly scaffolded, such as is common when students visit web sites. Indeed, this level of interest is taken even further in studies directed towards understanding behavior in virtual learning environments [10], but for the purpose of the work reported here, emphasis will remain on more traditional web-delivered information.

Considering the use of eye-tracking studies to look at web site usability, there are several previously reported studies of note. First, it's important to note that some methods related to eye-tracking can be powerful, and powerfully misleading if incorrectly applied. Thus, Bojko argues [11] that heatmaps of behaviors such as fixation counts of subjects looking at the site, must be carefully defined and methodologies chosen wisely to meet the desired level of data reliability and utility. Similarly, Siirtola and Räihä [12], propose data analysis methods for eye-tracking data that reduces the voluminous data produced by such studies into categories of (1) fixation heatmaps, (2) time spent in areas-of-interest (AOIs) and (3) transitions between AOIs. Results from previous eye-tracking usability studies have emphasized the interaction of users with navigation capabilities [13]; the visual appeal and structure of design elements of the site [14]; and the role of simultaneous think-aloud protocols as a research methodology [15]. All of these studies find that the data provided from eye-tracking can provide important insight into how usability is influenced by various elements commonly found in web environments.

It is also important to recognize that within an academic environment, the means by which students are encouraged to obtain information from the web can play a role in how they interact with the resources there. Considerable interest has accrued to the question of student motivation in on-line courses where retention rates are often low [16, 17]. Recently, attempts to formalize student behaviors in terms of cognitive theory have suggested factors that may influence student persistence [18]. Importantly, the nature of the motivational drivers, internal or external, appear to play a key role in how well students learn in on-line environments [19]. Thus, there have been a variety of studies that provide information about how web site usability can be considered. In the work reported here a combination of usability and eye-tracking methodologies are applied to a web site on chemistry concepts related to climate change.

3 Methodology

The web-site studied in this work was developed by a collaboration of scientists associated with several scientific societies. The site URL is www.explainingclimatechange.ca and it was developed by faculty and students associated with the King's Center for Visualization in Science at The King's University College in Edmonton, Alberta. Ultimately, the science content is broken into nine lessons, most of which include interactive elements in addition to didactic textual material. There is an initial page on the site that provides overall introductory information, followed by navigation page where the nine lessons are enumerated and can be accessed by clicking on the appropriate icon/link. While the level of information contained in each lesson is broad, this navigation page also has mouse-over information indicating more specifics about the contents within each lesson.

The aim of this laboratory experiment was to study (1) the ease of understanding the layout of the website including its functions and content; (2) the simplicity of use of the website; and (3) the speed and accuracy of locating needed information. Eye-tracking measures were predicted to reveal that most or all students would exhibit similar patterns of interaction. In particular, it was expected that students would take more time to understand the operation of interactive tools on the site, such as the Earth's Atmosphere learning tool.

3.1 Participants

Eleven students who participated in this study were recruited from a chemistry course designed for non-science majors at a large Midwestern University. The course includes formal study of key chemical concepts in climate change, such as the greenhouse effect of atmospheric gases. Thus, the participants in this study were aware of fundamental science, but did not have a particular penchant for studying scientific topics. Subjects received extra credit for their class and provided informed consent. Due to poor eye calibration, one participant was excluded, thus resulting in n=10 participants (5 males and 5 females; 4 freshmen, 3 sophomores, 1 junior, and 2 seniors; age mean age = 20.4.)

3.2 Tools and Materials

During the study, both the researcher and the participant were present in the same room. The researcher's computer was separated by a divider from the participant's computer and monitored the participant's eye-movement and task progress, allowing researcher to take notes and provide assistance when needed. Eye movements were recorded with a VT2 Eye-tracker [20]. The hardware for this system was located below the participant's computer screen and unobtrusively captured the respondent's eye movements. The accuracy of eye fixation was +/-0.5 degree (0.5 degree of visual angle amounts to approximately 0.5 cm on the screen at a 50 cm distance). This level of accuracy of recordings was satisfactory for the inferences about usability patterns made here. Eye-tracking data were recorded with iMotions software [21].

Another tool that was incorporated in the design of this project was UserZoom [22] - an online usability testing tool that allowed several advantages including; (1) the randomization of questions; (2) posting of instructions for participants to reference at the bottom of the screen; (3) time management for each task – timing out students when a pre-set limit was reached; and; (4) participant navigation through the tasks. Figure 1 illustrates how UserZoom was used and intergraded in our design.

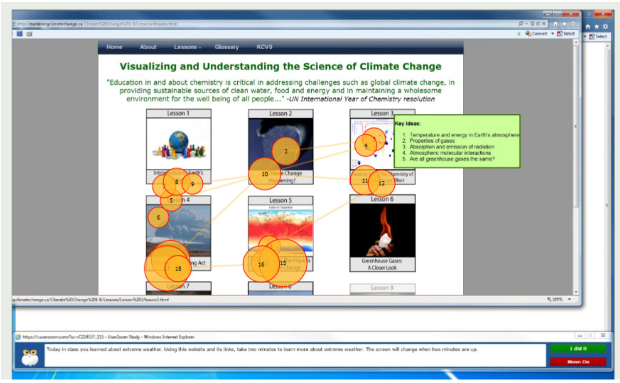

Fig. 1. A gaze plot of one user's fixations. This participant looked at the image and lesson title areas. The bottom window depicts how instructions were presented during the task. Participants were provided an opportunity to move to the next task by clicking on "I did it" button signaling task completion or "Move On" button signaling task fatigue.

By itself, eye-tracking is only able to determine where users look. A pure count of fixations cannot tell researcher whether users are productive, happy, or confused when they look at some elements of the website and ignore others. Thus, to augment the data from eye-tracking, both quantitative studies (such as measuring time spent on each task or the average time per task for a participant) and qualitative studies via personal interviews, were used to collect additional insights.

3.3 System Usability Scale

There are numerous instruments available to assess the usability of a website. One of the most robust and low-cost tools in evaluating the interface types is the System Usability Scale (SUS). Multiple studies confirm that the SUS is highly reliable (α = 0.91) and can be applied in evaluating the interfaces of a variety of products and services including websites [23]. It consists of ten equally balanced questions that ask participant to indicate the degree of agreement or disagreement with the statement on a 5-point scale ranging from *Strongly Disagree* to *Strongly Agree*. SUS is generally used shortly after the subject has interacted with the product, but before discussing it with the researcher. Instructions given to participants emphasize that they complete all items and mark the first answer that stands out rather than think about questions for a long time.

SUS score ranges between 0 and 100, 90 and above indicating that a product is exceptional, the score ranging in the 80s – good, and in 70s - acceptable [24, 25]. Negatively worded statements are reversed scored in obtaining the overall SUS score reported here. Initially responses are recorded as ranging from 1 to 5, but for positively worded items, a value of 1 is subtracted from the response value so they range from 0 to 4. For negatively worded items the user response value is subtracted from 5. Finally, all responses are summed and that value is multiplied by 2.5 to achieve the 100 point scale [24].

3.4 Task Procedures

All participants performed five tasks: (1) explore the website; (2) learn more about extreme weather; (3) learn when chlorofluorcarbons were invented; (4) locate and use Earth's Atmosphere learning tool; and (5) explore the definition of ocean acidification. The time allowed for the completion of any task was intentionally quite limited, varying from 2 to 5 minutes depending on level of interaction as judged by results from pilot study efforts using chemistry graduate students who were expected to be more familiar with the basic science associated with the site. Before proceeding with the task, participants were given the instructions (what the participant needed to do and how much time was allocated to the task.) Only after clicking the "Start" button, was the participant presented with the main menu page of the site (which occurs after an introductory page) and could start working on the task. Throughout the task, the instructions were available at the bottom of the page so participant could revisit them, if needed (Fig.1).

When subjects arrived, they were first introduced to the study and asked to sign consent forms. Then the experimenter gave a quick overview of eye-tracking system and asked participants to complete a short eye-calibration exercise followed by a brief demographics survey. After the survey, participants were instructed to complete five tasks. For each participant the first task was exploratory so that the participant would be able to become more familiar with the basic ideas of the website within two minutes. Participants were then introduced to four tasks that were presented in random order. After the experiment participants were asked to complete online System Usability Scale survey and to participate in a brief interview with the researcher. The interviews were audiotaped and included the video of desktop activity. All sessions were conducted on an individual basis and lasted between 50 and 60 minutes.

4 Results

4.1 Site Navigation and Eye-Tracking

A number of usability variables were measured including; the ease with which the site layout was understood by participants; the perceived simplicity of the site; the accuracy with which participants completed the tasks; and the speed of task completion. (Speed of task completion is derived from three quantitative measures that are often employed in usability work: average task time for both success and non-success groups, success score, and the number of clicks).

Table 1. Summary of success/non-success. Exploratory task #1 is not included. Number in each column represents number of participants. Number in braces indicates the mean time spent on the task. Numbers in brackets indicate information on clicks [min-max; mean].

Task	**Participant indicated actions**			**Predetermined metric**	
	Task success	Task non-success	Abandoned	Task success	Task non-success
2	3 {1:31} [2-11; 6]	7 {2:00} [9]	0	5	5
3	1 {0:59} [2-2; 2]	9 {2:00} [16]	0	5	5
4	3 {2:34} [0-10; 5]	6 {5:00} [18]	1 {0:38} [2]	4	5
5	1 {2:39} [14-14; 14]	9 {3:00} [12]	0	0	10

Table 1 shows a summary of the success or non-success on the final four tasks by all participants. (Task 1 is not included as it was strictly to explore the site, and all participants were able to do so.) Participants indicated success by clicking on a button “I did it”. As depicted, the success indicated by participants did not match the actual results, as determined when the task was constructed. For example, while one participant has indicated that he was successful on Task 3 (when were CFCs invented) completion, a total of 5 participants were successful at this task, in that they arrived at the page where the information was available. This data suggests that while participants are able to navigate the site reasonably well, they are not always certain they have found the correct information in the short time frames allowed in this study. Note that the site has a relatively large amount of textual information, so when students arrive at the correct location, they still need to read material, so they may not have had time to click the “completed” button despite arriving at the page with the needed information. Overall, the non-success rate is still somewhat high (50% for 3 tasks and 100% for 1 task), so the usability for finding specific information about content related to the chemistry of climate change appears to be subject to issues of the grain size of the information content in the navigation aides relative to the tasks assigned in this study.

Another way to assess student navigation of the web site is to analyze the pages visited or “clicked through”. For example, in task 2 students were asked to learn more about extreme weather as an impact of climate change using this website and its links and given two minutes. Participants exhibited little difficulty locating the correct initial lesson page (in this case Lesson 5), and interviews confirm that the mouse over information on the lessons navigation page assist significantly in this component of the task completion. The click through data for this task revealed that on average participants spent 1:31 minutes on the task (st. dev. 0:13 min), minimum time = 1:15 minutes. Three students indicated successful completion by moving to the next task. However, two students from non-success group have also completed this task successfully, in that they arrived at the specific page (within lesson 5) that describes the issue of extreme weather. Figure 2 illustrates the diagram of click path for Task 2. It demonstrates that 70% of participants visited the correct initial page - Lesson 5. Nonetheless only 30% of participants who visited Lesson 5 reported success (highlighted in green color) before time ran out. 30% of all participants visited, however, incorrect pages (highlighted in yellow) and showed relatively limited amount of exploration.

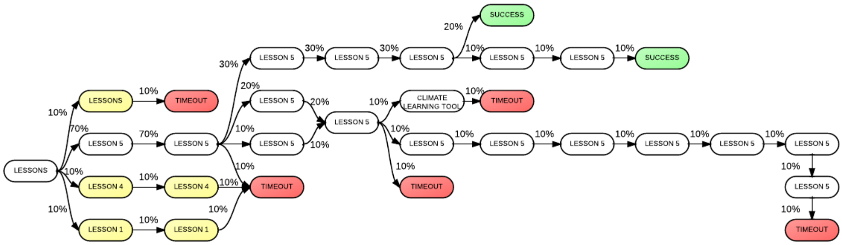

Fig. 2. Diagram of path measured in clicks for Task 2. Blocks highlighted in red show timeout, green blocks indicate success rate, and yellow blocks indicate incorrect route.

For the navigation oriented tasks described thus far, eye-tracking information tends to corroborate that students are using the intended navigation features of the site. The place where eye-tracking reveals important additional information is Task 4. In this task, students were asked to interact with an interactive feature on the site. This applet allows the participants to move an object that looks like a weather balloon up and down to explore features of the atmosphere, such as average temperature as a function of altitude. Directions for the use of the tool are presented on an introductory page and were not present on the page where the tool actually resided. In addition, designers of the tool used animations of meteors falling in the upper parts of the atmosphere or a plane flying through the sky. These animations occur at random times and as might be expected, eye-tracking shows that the movement strongly attracts student attention when they are using the tool. As a result, in general students' performance on Task 4 was affected by animated distracters. Figure 3 shows the eye gazing for participants 6 and 10.

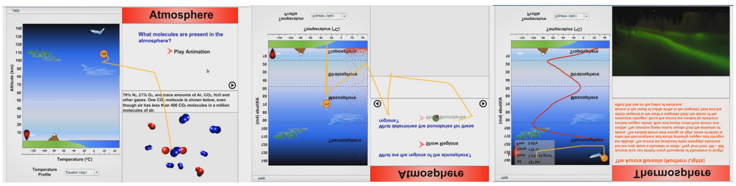

Fig. 3. A closer look at the images on the Task 4 page. Moving objects such as flying plane and falling meteors distracted participants from completing the task.

4.2 Participant Impressions of Usability

The key quantitative tool used to measure overall impressions of the site was the SUS. The calculated average SUS score over the 10 participants was 77.75 which indicates that users find the usability acceptable, but that there are ways that usability could be improved.

To hone in on ways to consider improving usability, interviews provide key quantitative data about users' subjective perspective of the overall site. During the interview, participants were asked to indicate what changes in design they would suggest. The top recommendation that 8 out of 10 students mentioned was to re-design the "green boxes" referring to the mouse-over boxes on the lessons navigation page (Fig.1). These mouse-overs were informational only, so navigation was in some sense directed by them, but not actually facilitated by them. Participants have suggested adding the links to content within the mouse-overs would dramatically improve navigation and allow the user to reach a desired page with just one click.

Because participants were asked to perform a variety of search activities, 7 out 10 suggested during the interview that including a search tool to the toolbar would be an important navigational feature. This feature would allow a user to faster locate the page that contains needed information.

5 Conclusions

This study looked at an academic web site with a relatively high level of information, including extensive text and interactive tools for understanding the chemistry associated with climate change, an arguably complex scientific topic. The tasks utilized in the study were devised to mimic activities that one might anticipate a teacher would ask a student to do given access to this site, or that a student who finds this site while doing self-directed research, might be trying to find. The tasks were intentionally designed to be short in time duration because there are many sites with information on climate change, and typical user behavior would be to move to a different site if information is difficulty to find or seems to take the user too long. The usability data tend to suggest that students are able to navigate easily through the website and find needed content relatively quickly. Thus, the main navigational features of the site, in particular the division of information into nine "lessons" appears to mesh with student-user experience for finding information. Interactive tools, not surprisingly, result in more detailed user interaction than tasks that involve navigating through text-based information to find answers for assigned tasks. Indeed, eye movement analysis suggests that details about interactive learning tools, such as animations that are added to provide an aesthetic component but do not advance the content information per se, tend to present a significant distraction for users. In the case noted here, rather than focusing on the task, students paid more attention following a shooting star or a flying plane. Thus, designers of academic sites should consider carefully decisions made to improve aesthetic impressions to decide if the gains from the added element are worth more than the possible distractions they can cause. Additionally, minor changes in design to existing navigational aids such as adding the links within mouse-over information and a search tool, can considerably improve the usability of an academic site.

Some of the basic results of this usability study have already been incorporated into improvements on this site. For example, the animations that provided distractions in the interactive tool have been removed. The usability study will be carried out again to determine if these adjustments and other improvements in navigation (such as replacing the word “glossary” with “definitions” on the glossary link) will enhance the SUS score from users. The generalizability of these observations will be further tested as well, by carrying out usability studies with other sites about this topic to see how variation in navigational aids and other features lead to different impressions by users and different success rates for specific tasks, such as those explored here.

Acknowledgments. Funding for this project was provided by the National Science Foundation under grant DUE-102292. Assistance from J. Raker, and K. Linenberger including helpful comments and pilot testing of tasks assigned to participants is gratefully acknowledged. Technical support for the use of the eye-tracking system from A. Peer was instrumental to the progress of this project.

References

1. Smith, S.D., Caruso, J.B.: The ECAR Study of Undergraduate Students and Information Technology. Educause, Louisville (2010)
2. Rosenfeld, L., Morville, P.: Information architecture for the World Wide Web. O’Reilly, Campbridge (1998)
3. Jacob, E.K., Loehrlein, A.: Information Architecture. Ann. Rev. Inform. Sci. & Tech. 43, 147–186 (2009)
4. Boatright-Horowitz, S.L., Langley, B.A., Gunnip, M.: Depth-of-Processing Effects as College Students use Academic Advising Web Sites. Cyberpsychology & Behavior 12, 331–335 (2009)
5. Fang, X., Holsapple, C.W.: An empirical study of web site navigation structures’ impacts on web site usability. Decision Support Systems 43, 476–491 (2007)
6. van Schaik, P., Ling, J.: The role of context in perceptions of the aesthetics of web pages over time. Int. J. Human-Computer Stud. 67, 79–89 (2009)
7. Tuch, A.N., Roth, S.P., Hornbaek, K., Opwis, K., Bargas-Avila, J.A.: Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI. Comp. in Human Behav. 28, 1596–1607 (2012)
8. Schonborn, K.J., Anderson, T.R.: A Model of Factors Determining Students’ Ability to Interpret External Representations in Biochemistry. Int. J. Sci. Educ. 31, 193–232 (2009)
9. Kramer, I.M., Dahmani, H., Delouche, P., Bidabe, M., Schneeberger, P.: Education Catching Up with Science: Preparing Students for Three-Dimensional Literacy in Cell Biology. CBE-Life Sci. Educ. 11, 437–447 (2012)
10. Persky, S., Kaphingst, K.A., McCall, C., Lachance, C., Beall, A.C., Blascovich, J.: Presence Relates to Distinct Outcomes in Two Virtual Environments Employing Different Learning Modalities. Cyber Psychology & Behavior 12, 263–268 (2009)
11. Bojko, A(A.): Informative or Misleading? Heatmaps Deconstructed. In: Jacko, J.A. (ed.) HCI International 2009, Part I. LNCS, vol. 5610, pp. 30–39. Springer, Heidelberg (2009)

12. Siirtola, H., Räihä, K.-J.: Using Gaze Data in Evaluating Interactive Visualizations. In: Ebert, A., Dix, A., Gershon, N.D., Pohl, M. (eds.) HCIV (INTERACT) 2009. LNCS, vol. 6431, pp. 127–141. Springer, Heidelberg (2011)
13. Leuthold, S., Schmutz, P., Bargas-Avila, J.A., Tuch, A.N., Opwis, K.: Vertical versus dynamic menus on the world wide web: Eye tracking study measuring the influence of menu design and task complexity on user performance and subjective preference. Comp. Human Behav. 27, 459–472 (2011)
14. Djamasbi, S., Siegel, M., Tullis, T.: Generation Y, web design, and eye tracking. Int. J. Human-Computer Studies 68, 307–323 (2010)
15. Gerjets, P., Kammerer, Y., Werner, B.: Measuring spontaneous and instructed evaluation processes during Web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. Learning & Instruction 21, 220–231 (2011)
16. Morris, L.V., Finnegan, C., Wu, S.-S.: Tracking student behavior, persistence, and achievement in online courses. Internet and Higher Education 8, 22–231 (2005)
17. Hart, C.: Factors Associated with Student Persistence in an Online Program of Study: A Review of the Literature. J. Interactive Online Learning 11, 19–42 (2012)
18. Lee, M.-C.: Explaining and predicting users' continuance intention toward e-learning: An extension of the expectation-confirmation model. Comp. & Educ. 54, 50–516 (2010)
19. Lee, M.K.O., Cheung, C.M.K., Chen, Z.: Acceptance of Internet-based learning medium: the role of extrinsic and intrinsic motivation. Information & Management 42, 1095–1104 (2005)
20. Eyetech, http://www.eyetechds.com/research/vt2
21. iMotions, http://imotionsglobal.com/
22. UserZoom, http://www.userzoom.com/
23. Brooke, J.: SUS: A "quick and dirty" usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, I.L. (eds.) Usability evaluation in industry, pp. 189–194. Taylor & Francis, London (1996)
24. Bangor, A., Kortum, P., Miller, J.: Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of Usability Studies 4, 114–123 (2009)