Evaluating the Effectiveness of Interactive Map Interface Designs: A Case Study Integrating Usability Metrics with Eye-movement Analysis

Arzu Coltekin, Benedikt Heil, Simone Garlandini, and Sara Irina Fabrikant

ABSTRACT: This paper proposes combining traditional usability methods with the analysis of eye movement recordings to evaluate interactive map interfaces, and presents a case study in support of this approach. The case study evaluates two informationally equivalent, but differently designed online interactive map interfaces presented to novice users. In a mixed factorial experiment, thirty participants were asked to solve three typical map-use tasks using one of the two interfaces; we then measured user satisfaction, efficiency (completion time) and effectiveness (accuracy) with standard SEE usability metrics. While traditional (bottom line) usability metrics can reveal a range of usability problems, they may be enhanced by additional procedural measures such as eye movement recordings. Eye movements have been shown to help reveal the amount of cognitive processing a display requires and where these cognitive resources are required. Therefore, we can establish how a display may or may not facilitate task completion by analyzing eye movement recordings. User satisfaction information related to tested stimuli (i.e., collected through standardized questionnaires) can also be linked to eye tracking data for further analysis. We hope that the presented methodology and case study will help cartographers and map interface designers to better identify design issues in their products, and that these insights will eventually lead to more effective and efficient online map interfaces.

KEYWORDS: Geographic visualization, human computer interaction, eye movement analysis, interactive maps, interface evaluation

Introduction

About ten years ago, Howard and MacEachren (1996) predicted that digital "softcopy" maps were becoming the norm replacing paper maps, and as a result, the design of interface tools would become as fundamental to cartography as the design of the maps themselves. The spread of high-bandwidth Internet and access to it through the increasing use of location-based services with mobile devices (i.e., in-car navigation systems, personal digital assistants, cell phones, etc.) seems to have validated this prediction. In accordance with this development, new complex representation forms and interactive methods for visualizing geospatial data are available to large audiences with different levels of experience in handling them (Koua et al. 2006; Fabrikant et al. 2008). The need to assess the impact, usefulness, and usability of these tools is increasing at the same rate as their rising availability and spreading versatility (Fuhrman et al. 2005; Koua et al. 2006; Nivala et al. 2008; Haklay and Zafiri 2008). However, as identified by MacEachren and Kraak (2001) already at the turn of the century, evaluating highly interactive visual map interfaces in a cognitively informed setting requires established paradigms; the geovisualization community is still faced with this challenge. We hope to contribute to the efforts for tackling this issue by bringing modern usability engineering techniques together with eye movement analysis.

Arzu Çöltekin, **Benedikt Heil**, **Simone Garlandini**, and **Sara Irina Fabrikant**, Dept. of Geography, University of Zurich, Winterthurerstr. 190, 8057 Zurich, Switzerland. E-mail: <{arzu.coltekin, simone.garlandini, benedikt.heil, sara.fabrikant}@geo.uzh.ch].

Usability Engineering

Usability engineering refers to a set of techniques and concepts for assessing a product or a system's ease of use based on systematic evaluations, system inspection, and inquiry methods (Good et al. 1986; Nielsen 1993). Typically users are provided with a specific set of tasks based on a particular usage scenario, and in a specific context. Usability performance metrics such as satisfaction, efficiency and effectiveness (SEE) are employed to assess how easy the product or system is to use. Satisfaction refers to a user's attitude or preferences about the system, efficiency refers to how quickly the tasks are completed, and effectiveness refers to whether or not a task is successfully completed.

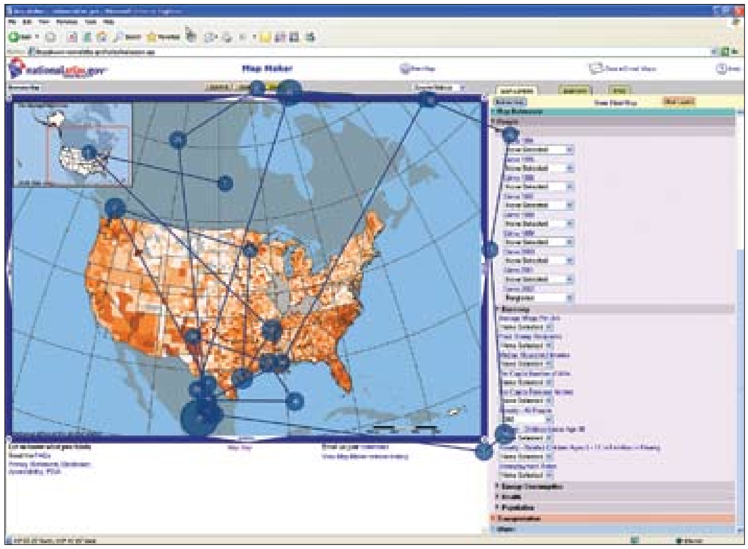

Figure 1. A scan-path plot showing one participant's fixations and saccades over a selected time period. When fixation duration is longer, the circle that represents this fixation is represented proportionally larger.

Iterative evaluation sessions allow usability researchers to identify most usability problems (Nielsen 1993). In typical usability studies, human–system interactions are evaluated with direct observation, pen and pencil questionnaires, video analysis, or key stroke and mouse click recordings. These evaluation procedures are often supported with other standard empirical methods such as think-aloud protocols and interviews. When the procedure involves explicit self-reports or interviews, various psychological and social factors can influence human behavior (and performance) and thus create bias in the results, such as short-term memory problems, anxiety or desire to "succeed" in a test situation.

With self reports or interviews, what people say or believe they do, is not always what they actually do. This can be particularly relevant in highly interactive systems used to solve complex problems, and when people might not be able to fully verbalize their own complex inference making. A viewer's *cognitive load* might become so high during task completion that verbal reports or think-aloud protocols interfere with the quality of inference making. Eye movement recordings, on the other hand, can offer additional unobtrusive evidence of overt user behavior. Eye movement recordings are frequently viewed as a window into internal cognitive processes (Bojko 2006; Goldberg et al. 2002). By studying them, we may be able to compensate for the excessive cognitive load that prevents the participant from remembering processes when self-reporting. Eye movement recordings are also very useful for identifying *where* problem areas are in system use and *how* the information might be processed.

In geographic information science (GIScience) literature there have been many usability studies. Some earlier examples in the digital era include Medyckyj-Scott (1993), Nyerges (1993), Knapp (1995) and more recent examples such as Fuhrmann et al. (2005), Robinson et al. (2005), and Harrower and Sheesley (2005). A recent special issue of the *Cartographic Journal* (volume 2, issue 45) is entirely devoted to "Use and User Issues in Geographic Information Processing and Dissemination," where we see relevant work on various topics. Within the cartography and geovisualization domain, traditional (static) map display evaluation methods also have been based on standard usability approaches (i.e., testing, inspection, and inquiry methods). Recent research papers frequently point at the need for better, more suitable methods for evaluating maps, stating that the standard methods "may no longer be suitable for the growing range of map users, usage scenarios, and digital map devices" (Nivala 2008), in particular with new interactive visualizations (Koua and Kraak 2004).

In this study, we aim to perform cognitively informed research and to explore the potential of eye tracking in this domain. Therefore, similarly to Fabrikant et al. (2008), we propose a combination of traditional usability engineering methods with eye movement analysis for the empirical evaluation of interactive map interfaces.

Eye Movement Studies for Interface Evaluation and Usability

Eye tracking has a history of nearly one hundred years in psychology, but early technology was cumbersome, difficult to use, and prohibitively expensive (Dix et al. 2004; Duchowski

2007). Technological developments in recent decades have made eye tracking systems more accessible and available. Modern eye-tracking systems allow fairly accurate recordings of pupil diameter, number of fixations, fixation durations, and saccades multiple times per second during a session. A fixation is when eyes are stationary during a given threshold of approximately 50 to 500 milliseconds (reported in Irwin 2004; Henderson and Ferreira, 2004 and Bojko 2006) and a saccade is the rapid eye movement that occurs between fixations. A representation of fixations and saccades can be seen in Figure 1. Analysis of eye movement recordings has been increasingly and successfully employed in various fields such as software design and interactive web interface evaluation research and practice (e.g., Goldberg and Kotval 1999; Byrne et al. 1999).

When utilizing eye movement analysis to evaluate the usability of an interface, some common assumptions are that more fixations may indicate a less efficient search strategy, longer fixations may indicate difficulty with the display, and plotting scan paths and fixations will allow documenting what people look at, how often, and how long (Goldberg and Kotval 1999; Bojko 2006). When users are searching to find the correct link, button, or another control on an online interface, typically two types of processes occur: a perceptual one (where the user should locate/notice the target) and a cognitive one (where user cognitively computes the visual input and understands the function of the target). Eye movement analysis provides valuable quantitative and qualitative information on both stages of visual search and thereby complements SEE metrics (Goldberg and Kotval 1999; Jacob and Karn 2003). These observations have led some recent academic and industrial interface evaluation studies to combine eye movement analysis with other usability methods (e.g., Pretorius et al. 2005; Bojko 2006).

Evaluating Interactive Map Interfaces with Usability Engineering Methods and Eye Movement Analysis

Utilizing information that can be gathered by recording eye movements to understand the relationship between map reading and map design was reported as early as the 1970s (Steinke 1987). The cartographic community showed interest in eye tracking until the 1980s, but after this decade, the interest seems to have nearly disappeared (Steinke 1987; Brodersen et al. 2002; Fabrikant et al. 2008). This trend may be a result of a suboptimal cost-benefit relationship; eye movement analysis was financially costly to start and effort-intensive to finish. Today, eye tracking hardware is affordable and even though analyzing eye movement data still is a time-consuming and complex process, digital processing can arguably make it easier to process very large datasets in comparison with the analog methods used in the 1970s and 1980s.

Maps have also changed since the 1980s. Interactive digital maps (as opposed to static paper maps) have become more complex to analyze due to added dynamic features. Digital interactive map interfaces typically come with two display elements: a cartographic data display area, where the map itself is presented, and a set of graphical user interface (GUI) elements which allow for interaction with the presented map data. The usability of such maps relies heavily on interface design (You et al. 2007). By employing the eye movement data collection method for complex interactive map interfaces, we can monitor a user's inference-making process while interacting both with the map and the interface elements at the same time (Fabrikant et al. 2008). The questions *where*, *when*, *how long*, *how often,* and *in which order* a display element was attended to during a task may allow us to interpret more effectively why task completion or inference making might be facilitated (or hindered) with a particular interface design, and whether the map interface is indeed utilized as intended by the designers. Procedural (eye movement) data combined with baseline effectiveness and efficiency data (i.e., accuracy and speed of response) provides added value to the process of systematically evaluating interactive map interfaces.

Experiment

The proposed evaluation methodology has been applied to a controlled experiment comparing two interactive online map interfaces: the *Map Maker* service of the *National Atlas of the U.S.A.* (Natlas 2008), and an interactive thematic map published on the *carto.net* web site (Carto.net 2008). Participants were asked to perform a set of map-use tasks while their eye and mouse movements were being recorded. Although the maps include the same statistical data, they differ significantly in the approach taken to map interface design (Figure 2).

While using these two representations, users were consistently able to answer our three experimental questions (these questions are introduced in the

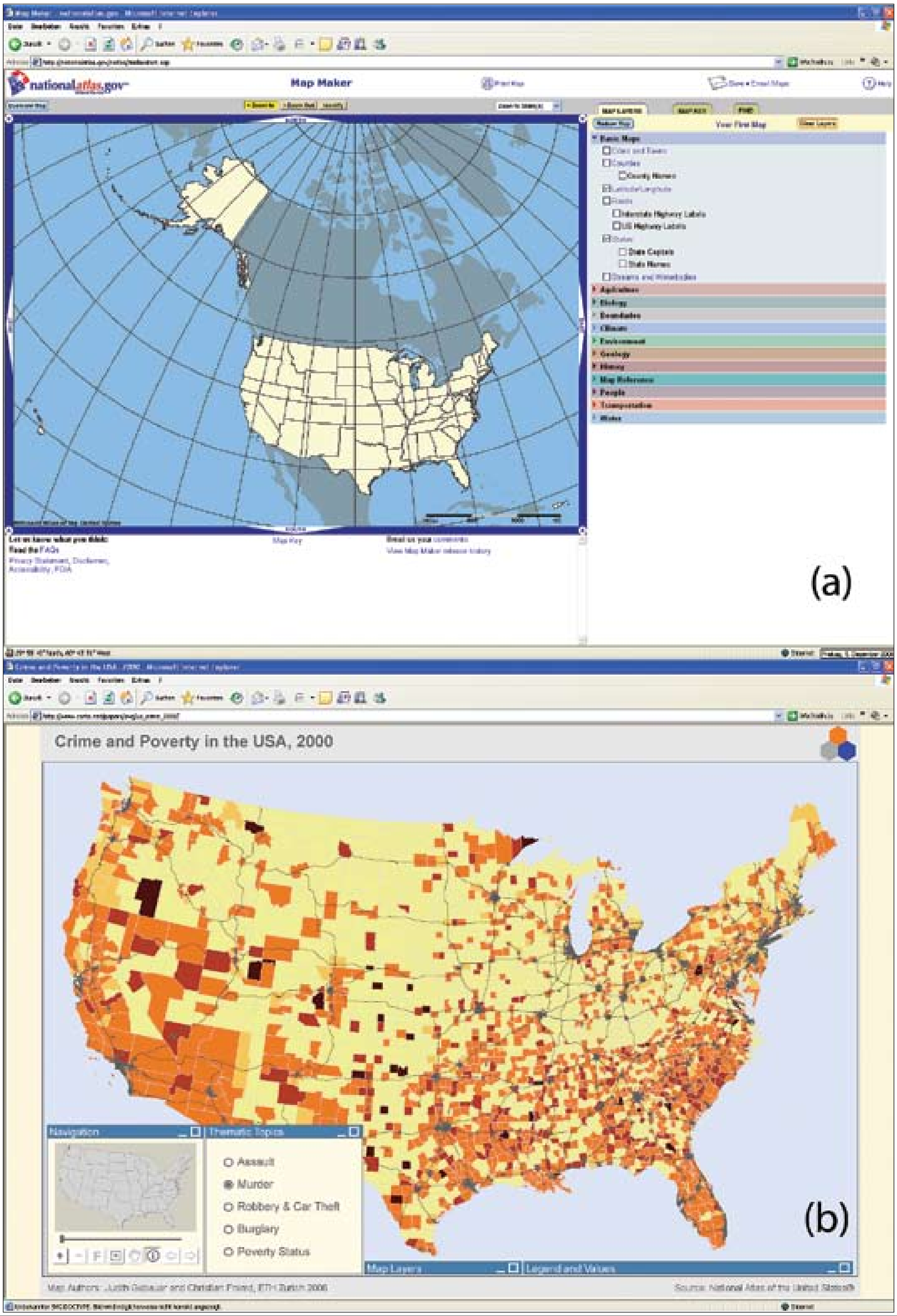

Figure 2. Screen shots of two interactive maps used in the study: National Atlas of the United States (Natlas) (a), Carto.net (b) showing crime and poverty data taken from the Natlas.

section Experimental Design). In other words, in the frame of our experiment, two representations are informationally equivalent, however, we contend that they are not computationally equivalent (Larkin and Simon 1987). In Larkin and Simon's words, the definitions of these two concepts are as follows: "Two representations are informationally equivalent if all the information in one is also inferable from the other, and vice versa. Each could be constructed from the information in the other. Two representations are computationally equivalent if they are informationally equivalent and, in addition, any inference that can be drawn easily and quickly from the information given explicitly in one can also be drawn easily and quickly from the information given explicitly in the other, and vice versa." (Larkin and Simon 1987, p. 66).

Based on Larkin and Simon's computational equivalence concept we hypothesize that map users will perform less efficiently with Natlas as it requires comprehension of a more complex interface.

Designer Interviews

To better understand the design process and usage contexts, we first studied the stimuli by documenting both the technology that was involved in making them and by interviewing the individuals who were involved in decision-making roles of the development and design process. Display designers responded to a ten-question online form inspired by the system usability scale (SUS) questionnaire (Brooke 1996), a standard usability measurement tool typically used to measure a user's attitudes and system preferences. Designers' answers to the re-engineered SUS questionnaire were later employed as a baseline to compare actual user responses to their SUS questionnaire answers (see section: System Usability Scale and Participant Interviews).

The interview revealed that *Natlas* was developed using ESRI's Map Objects Internet Map Services to render the maps and ArcIMS to manage the communication between the Web server and five spatial servers. The graphical user interface (GUI) has evolved through several development environments, all of which are still present. They include HTML, JavaScript, Cold Fusion, and Active Server pages. There are also tables for zip codes and geographic names which are managed by an Oracle database on a Sun server (UNIX). The system does not require a special plug-in, and it runs inside mainstream web browsers.

Carto.net designers, on the other hand, developed the interface based on Scalable Vector Graphics (SVG) to render the graphics, and ECMAScript (*European Computer Manufacturers Association* script) to handle map interaction. Only Internet Explorer users need a specific SVG plug-in (i.e., the Adobe SVG viewer). All other browsers can display the map without any additional installations.

The intended audience for both maps is "the average internet user," that is, non-domain experts, without any specific additional technical expertise. For both maps, the design team included at least one cartographer. *Natlas* required a two-year design and development period, with 12 people involved at different stages over the entire development time. Natlas was designed with frequent usage in mind. The system was thoroughly tested before public launch, including a classical usability study that led designers to "*deliberately remove[d] functionality that was too complex for our average users*." Currently, it is maintained by one employee and receives more than 150,000 unique visits per month.

Carto.net's implementation, on the other hand, grew out of a student's class project. Four people have been involved in the development and design process; two in supervising roles, and two implementing the system.

While developers of Natlas and Carto.net made different technical, cartographic, and interface design choices, the map data were taken from Natlas, and thus are identical. The *Carto.net* designer that we interviewed stated that "*the application is not too flexible and customizable*" and that it could be improved in these areas. The map interface was not designed with frequent usage in mind, and no usage statistics are available for this map. It was not subjected to any pilot testing or human subject testing, before it was launched. However, the designers' intention was that it "*should be easy to use for everyone.*"

Experimental Design

Our experiment was designed to provide a balance between experimental control and ecological validity. In a between-subject design, we monitored user responses (N = 30) to three different typical map-use tasks (independent variable). We used tasks that have different levels of complexities: two are close-ended questions and require an inference related to an attribute or a location (Questions 1 and 2), and one is an open-ended question that requires the participant to compare two spatial distributions (Question 3). Test questions were as follows:

1. What is the number of assaults in Washington County (Maine) in the year 2000?
2. Which county in the State of Oregon has the highest murder rate in the year 2000?
3. Looking at the map of the U.S.A, overall, do you see a relationship (if any) between poverty rates and burglaries in the year 2000?

The tasks were presented in a systematic rotation to counter-balance for a potential learning effect. The dependent variables include the traditional usability measures such as response time (efficiency measure) and accuracy of response (quality measure). These performance measures are complemented by self-reports collected using a standardized system usability scale, SUS (Brooke 1996). Additionally, eye movement recordings include gaze plots and fixation patterns in selected areas of interest (AOI) in the interface. These AOIs were selected based on a cognitive walkthrough session that was performed before the experiment, and they were confirmed or enhanced after the recording session, taking into account where participants reported having trouble. These gaze plots and fixation patterns in the selected AOIs allow us to link traditional usability (success) measures with users' interface interaction processes. Finally, participants also provided qualitative interface preference feedback. Even though the experimental design included two professional groups (geography-educated and others), we focused on identifying the usability problems based on map interface designs rather than on differences between the two groups.

Participants

Thirty people (11 females, 19 males) participated in this study. The average age was 28 years. Fifteen participants had college-level training in geography and fifteen participants had a nongeographic educational or professional background. All participants are nonnative, but fluent English speakers (the interfaces of both maps are in English). They have reported a high level of experience with the relevant operating system, the Internet, and the relevant browser. It is also important to note that the participants were, on average, fairly experienced in using graphics and spatial data. We asked our participants to rate their proficiency levels from 1 (no experience) to 5 (everyday use) using graphics of any kind (maps, charts, graphs, photos, etc.), in which geographers had an average score of 4.5, and non-geographers 3.8, and use of spatial data (maps, digital elevation models, remotely sensed images, etc.), in which geographers had an average score of 4, and non-geographers, 2. They were offered no compensation for their participation.

Materials

Two interactive online map interfaces (Figure 2) were selected as stimuli for the study (*i.e., Natlas* and *Carto.net* as introduced earlier). Both interfaces allow access to a dataset that represents thematic information related to "Crime and Poverty in the USA, 2000" and both provide several interactive features to display and query this dataset.

The online interactive mapping systems are based on two different interface designs: *Natlas* divides the screen into three main parts, where the top part of the screen has the title of the map and several interactive buttons and links. The map portion of the display takes about 43 percent of the screen, and the rest is reserved for interacting via buttons, links, and pull-down menus which are distributed in three distinct tabs (Map Layers, Map Key, and Find). When a user points at main buttons, Natlas provides a small window explaining the function of the button. The answer to a query is returned in a pop-up window.

Carto.net uses a larger area of the screen for displaying the map (72 percent) and four small windows are floating over on the map area. These windows can be minimized and/or moved. Queries are made via radio buttons. The query answer is returned in the top bar of the legend window as the mouse moves to the relevant area.

It is noteworthy that while *Carto.net* provides data and queries only for Crime and Poverty in the USA in 2000; Natlas has a very large selection of other themes as well as this dataset. To make sure that what was immediately visible on one interface was also immediately visible in the other, appropriate tab and pull-down menus were left open on both *Natlas* and *Carto.net* before the participants viewed the interfaces.

Setup

The experiment was performed on a Windows workstation, running the Tobii Studio software for automatic stimuli display and eye movement recordings. The SUS survey was delivered digitally via the Morae usability software. Interface stimuli were displayed on a 24-inch flat screen at a 1600 × 1200 screen resolution. Eye movements were recorded with a Tobii X120 eye tracker, at a 60Hz sampling resolution.

Procedure

After welcoming the participants, we requested them to sign a consent form that provided general information about the experiment. This was followed by a background questionnaire. Participants were then trained to locate the States of Oregon and Maine on a digital map of the conterminous U.S., where all other state

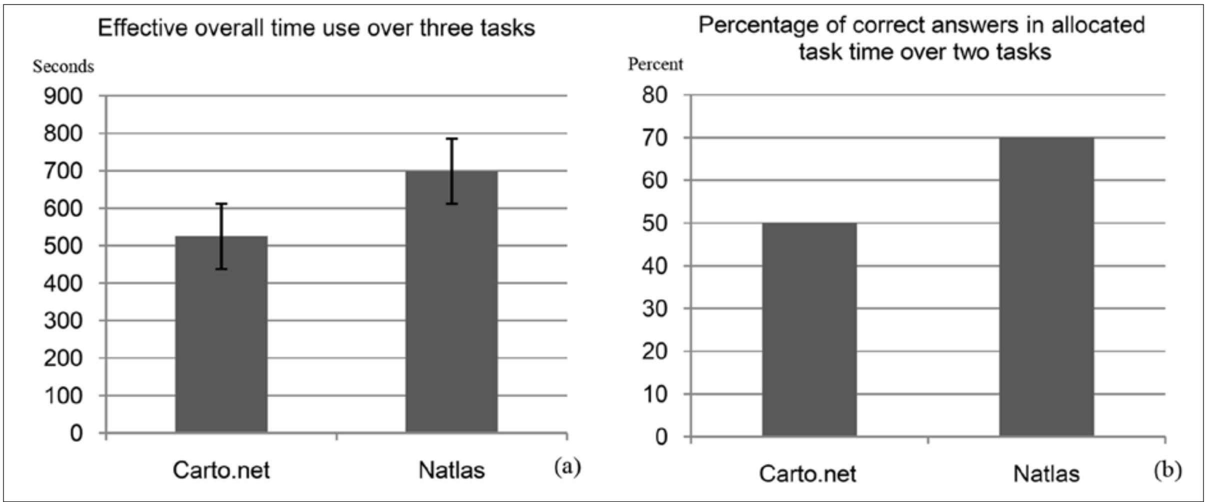

Figure 3. Overall mean response times with standard error (a) and accuracies for two map types (b).

names were removed, and the locations of these two states were highlighted. The purpose of the training was to make sure that participants knew where these states are located, as the goal of the experiment was not related to finding them on the map. In making sure that they knew where the states are located, we can be confident that their response time is related to solving the planned map-use tasks instead of trying to locate the states. Before recording began, participants were instructed to assume a comfortable position and not move too much, to maximize the eye movement recording's accuracy. Then, a calibration with the eye tracker followed. At this stage, participants were ready to start solving the tasks with the map interfaces and eye movement recording began. The experiment leader provided verbal instructions for carrying out the tasks during the experiment and participants also responded verbally. In order to limit the duration of the experiment, the experiment leader provided participants with help after five minutes, and if the participant got stuck, the task was considered incomplete. After completing the map-use tasks, participants filled out a close-ended feedback questionnaire and responded to three additional qualitative preference questions. After completing the questionnaire, participants were debriefed and thanked for their participation.

Results

All participants completed the three tasks using the two interfaces. Total completion times over all three tasks were: Carto.net minimum: 220.0s, maximum: 1305.0s, mean: 528.1s; Natlas minimum: 337.4s, maximum: 1120.9s, mean: 698.8s. This excluded technical problems (i.e., we needed to restart the browser in two sessions) and task delivery times. Following a common practice, we removed one participant in the efficiency evaluation because the measured value was several standard deviations away from the mean (e.g., Hegarty and Waller 2004).

Mean response times and accuracies are shown in Figure 3. To calculate their statistical significance, response time and accuracy scores were subjected to a one-way analysis of variance with map interface types as between subject independent variables. Overall, participants were significantly more efficient (faster) using *Carto.net*'s interface, $F = 7.359$, $p=.011<.05$, but significantly more effective (accurate) using *Natlas*, $F = 5.095$ and $p=.032<.05$.

Explored usability metrics regarding the response time and completion rate (accuracy) reveal an interesting problem for evaluating the interfaces: one of the designs allows users to perform faster while the other gives more accurate results. This tells us that both designs have elements that make the user perform better or worse in some ways, but why and how is this happening? This is where the eye movement analysis offers additional help by allowing us to study micro-level behaviors linked to people's visual attention and internal cognitive processes.

In Figure 4, participants' eye movement behavior is depicted with an interpolated fixation density surface overlaid onto the Natlas (Figure 4a) and Carto.net (Figure 4b) map interfaces.

For this study, the fixation filter values were set to a radius of 50 pixels, and the minimum fixation

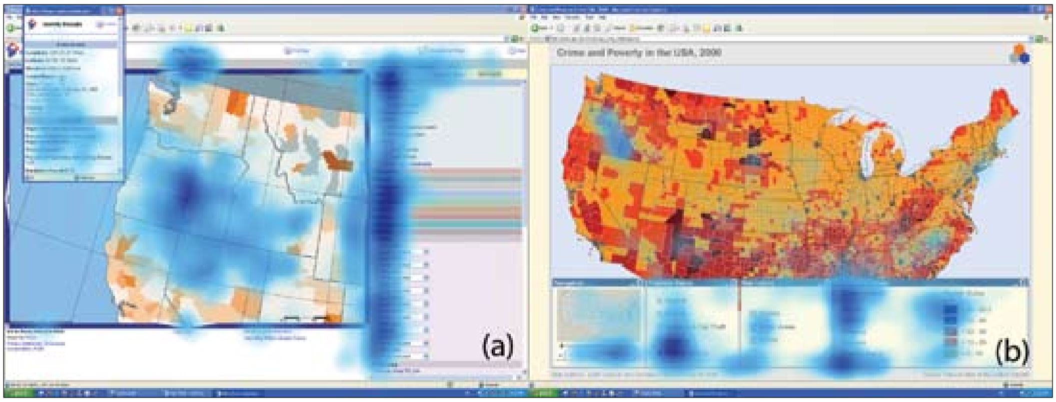

Figure 4. Density maps showing the aggregated gaze patterns over all tasks and all participants for both interfaces. (a) Natlas, (b) Carto.net.

duration (threshold) to 100ms. The fixation filter takes a series of gaze coordinates to be a single fixation if they stay within the given radius at the given threshold. Overall fixation counts (total number of fixations by all three participants for all questions) for each map were: Natlas (sum): 19554, Carto.net (sum): 17902. This implies that users might have had a less efficient search strategy with Natlas in comparison with Carto.net. Overall mean fixation durations for Natlas (Mean (M)= 7715.1s, Standard Deviation (SD): 139.8s) and Carto.net (M= 6642.1s, SD: 111.3s) also raise the question whether this might mean Natlas has a more complex interface. To analyze the sources of issues regarding search efficiency and difficulties we conducted a deeper level study of certain areas of interest.

Areas of Interest (AOI) Analysis and Identified Usability Issues

The areas of interest were defined before data collection began (Table 1(a) and Table 1(b)), based on which interface elements would have to be used to successfully solve the tasks. For this purpose, the authors performed a cognitive walk-through session before running any sessions, to determine which interface elements participants would likely use, and in what sequence. The identified interface elements and sequence were also cross-validated in a pilot experiment with a small set of test participants. Participant interviews also confirmed the anticipated problem areas.

Whole screen and *map area* on both maps were considered AOIs, and the identified areas of interest with potential usability issues were determined as *Identify* and *Redraw Map* buttons on Natlas and the *mouse roll-over behavior* that reflects on the *legend-bar* in Carto.net.

For Natlas, once the *Identify* and *Redraw Map* buttons were discovered—time to first fixation: *Identify* M = 116.2s, SD = 102.1s; *Redraw Map* M = 50.6s, SD = 30.6s) and their functions were understood (time to first mouse click: *Identify* M = 186s, *Redraw Map* M = 54s—, tasks were successfully completed. The difference between the mouse clicks and the first fixations tell us that the labeling of the *Redraw Map* button was more quickly understood by the users than that of *Identify* (i.e., this may indicate that the labeling of this button may benefit from further testing). Considering the average task completion time for the first task was 190 seconds for Natlas, participants spent 61 percent of their time to locate *Identify* and 27 percent for *Redraw Map*. The fact that there are 423 fixations before *Identify* and 302 fixations before *Redraw Map* buttons were located tells us participants were searching for them in other parts of the screen. Looking at the scan paths, we observe that the majority of the people (75 percent) spent time on the menu area (22 percent of the screen with tabs, located on the right side of the screen) looking for a tool that would help them.

The *Redraw Map* button has a salient feature: it flashes and pops up a small window telling the user to press the button. However, 30 percent of the users did not use this button but sought alternative ways to fulfill its function. Both buttons are probably too small: *Identify* 0.08 percent of the screen, *Redraw Map* 0.13 percent of the screen. It is also noteworthy that 100 percent of the participants who needed assistance (5 of 15 participants needed assistance to continue) within the allocated five-minute task duration limit had trouble with the *Identify* button (i.e., the experiment leader had to tell the users that they have to use the *Identify* button and showed

Carto.net AOI-Name	Screen percentage
Whole Interface	85.49% (1641408 pixels)
Map area	71.45% (1371840 pixels)
Zoom In	0.06% (1152 pixels)
Zoom Out	0.06% (1152 pixels)
Legend Value Default Position	8.57% (164544 pixels)
Legend Value First Half Default Position	3.34% (64128 pixels)
Legend Value Second Half Default Position	4.10% (78720 pixels)
Legend Value Infobar Closed	0.90% (17280 pixels)
Legend Values Infobar Open	0.90% (17280 pixels)

(a)

Natlas AOI-Name	Screen Percentage
Identify	0.08% (1536 pixels)
Redraw Map	0.13% (2496 pixels)
Zoom In	0.09% (1728 pixels)
Zoom Out	0.09% (1728 pixels)
Zoom To State	0.15% (2880 pixels)
Map Key	0.08% (1536 pixels)
Find	0.08% (1536 pixels)
Overview Map	0.07% (1344 pixels)
FAQ	0.05% (960 pixels)
Help	0.12% (2304 pixels)
Map area	42.79% (821568 pixels)
Pull-Down Menus	22.11% (424512 pixels)
Whole Interface	85.49% (1641408 pixels)

(b)

Table 1. Selected AOIs for (a) Carto.net and (b) Natlas.

them where this button was), confirming identified usability problems with this feature.

The reason why Carto.net is faster than Natlas might be partly due to the size and locations of these two critical buttons. However, we speculate that it has also to do with the map size: the map covers 43 percent of the screen for Natlas whereas it is 72 percent of the screen for Carto.net. This difference in size leads to less use of zoom buttons on Carto.net in comparison with Natlas. Both fixation counts and fixation durations suggest that participants have used zoom functions (for zooming in or out) less when using Carto.net.

While indicators regarding efficiency (speed) are positive for Carto.net, one usability issue that requires discussion regarding this interface manifests itself in the accuracy scores (task completion success). Carto.net presents the queried data as people roll their mouse over to the relevant geographic area on the bar above the legend. This proves to be counter intuitive for most users: 80 percent of the participants clicked on the map at least once and expected to see a result before they discovered this function; 40 percent of the participants used the right mouse click to explore what other options may be "hidden." This "mouse roll over" feature also resulted in 80% of the participants requiring assistance (it took approximately 51 seconds for an average participant to find out that the mouseover changed the values in the legend).

Average time to first fixation for the legend bar when it was closed was 43.5s, and time to first fixation for the legend bar when the window was open was 46.7s. The 3.2 seconds difference between these two values can be interpreted as the time taken for the participants to discover that this window had a dynamic behavior (i.e., that it can be opened). Five of fifteen participants needed assistance, and four of these five participants needed help with the use of the legend. The legend is designed to present information in two columns, which appears to mislead users. Overall, they spent more time looking at the first (left-hand) half of the legend (48 percent of the participants, observation length M = 24.5s, SD = 23.5s, fixation length M = 20.7s SD = 19.9s) than the second (right hand) half where the information is more relevant for two tasks out of the three tested (observation length M = 15.7s, SD = 19.6s, fixation length M = 13.3s SD = 17s). On the other hand, accuracy was high with Natlas interface because the responses are isolated within pop-up windows. Isolating the query results therefore reveals itself as a favorable design choice in this case.

An additional observation regarding Carto.net's design is that it is possible to move the information windows, yet only four users out of fifteen (26.6 percent) discovered that these windows could be moved. This is not relevant to a task in the experiment, however it is of note. Users could be made aware of this useful feature in an explicit manner. For Natlas an additional usability issue may be about the help button, as only seven percent of the participants fixated on this feature and only one participant used it.

System Usability Scale and Participant Interviews

There are a number of standardized usability surveys to measure participants' attitudes or

Modified System Usability Scale Question Rates 0: Strongly Disagree to 4: Strongly Agree	Carto.net (mean)	Natlas (mean)
Q1. I think I would like to use this interactive digital map frequently.	2.93	2.07
Q2. I found the interactive digital map unnecessarily complex.	2.86	2.73
Q3. I thought the interactive digital map was easy to use.	3.53	2.86
Q4. I think that I would need the support of a technical person to be able to use this interactive digital map.	3.73	2.80
Q5. I found the various functions in this interactive digital map were well integrated.	3.00	2.47
Q6. I thought there was too much inconsistency in this interactive digital map.	2.40	2.80
Q7. I would imagine that most people would learn to use this interactive digital map very quickly.	3.47	3.40
Q8. I found the interactive digital map very cumbersome to use.	2.33	2.07
Q9. I felt very confident using the interactive digital map.	3.13	2.87
Q10. I need to learn a lot of things before I could get going with this interactive digital map.	2.93	2.60

Carto.net SUS score: 63.33
Natlas SUS score : 54.17 (Where minimum is 0 and maximum is 100. Calculated according to Brooke, 1996)

Table 2. Modified system usability scale questions. These were rated by participants using a 5-step scale ranging from "Strongly Disagree" (0) to "Strongly Agree"(4).

preferences, such as the Questionnaire for User Interface Satisfaction (QUIS; Chin et al. 1988), the Computer System Usability Scale (CSQU; Lewis 1995), and the System Usability Scale (SUS; Brooke 1996). In a systematic comparison carried out by Tullis and Stetson (2004), the SUS, one of the simplest questionnaires studied, provided among the most reliable results across sample sizes. At the end of the eye movement recording session, participants filled in an interactive questionnaire prepared based on the System Usability Scale (SUS), including 10 Likert-style preference questions (Table 2) and 3 qualitative questions. As mentioned earlier, designer interviews before the experiment also included SUS-style questions so that designer intentions could be compared with user satisfaction.

Results revealed that Carto.net users had a more positive attitude towards the interface than did the Natlas users. This difference is not statistically significant in itself ($F=0.89$, $p = 0.354>.05$), however, when we look into the individual questions, it becomes statistically significant on two accounts: more people (question 1, $F = 4.95$, $p = 0.034<.05$) declared that they would like to use the Carto.net interface frequently, and more people (question 4: $F = 7.22$, $p=.012<.05$) declared that they felt they would need technical support when using the Natlas interface. While the original SUS questionnaire author (Brooke1996) suggests only the sum of the scores should be meaningful, we find this information revealing, especially in comparison with designers' expectations.

The results of this questionnaire were cross-checked against the designers' responses. On the two accounts that these were significantly different, for example, Natlas designers expressed that they expect the system to be used frequently by their users (with a score of 4), but the participants' average was only 2.1. The next question that yielded a significant difference between the two interfaces was regarding the expected technical skills of users. The designers report that they hope users with minimal or no technical skills can comfortably use the interface (Natlas=3, Carto.net=4). On the other hand, the participants' average reveals that a majority of them felt that they would need technical support to be able to successfully complete the tasks (Natlas=2.8 Carto.net=3.7).

In addition to 10 SUS-inspired questions where users ranked their preferences one to five, we added three additional qualitative questions:

1. Would you use this interactive digital map instead of a traditional Atlas? Please explain why.
2. Would you recommend this interactive digital map to a friend? Please explain why.
3. Which of the following features do you believe need major improvement?
4. Server and process speed / Map coloring scheme / Size of the buttons / Placement of the buttons / Wording of the buttons / All of the above / None of the above / Other (please tell us which)

Participants were asked to check one or more of the offered categories and/or add their own categories. Summarizing participants' open-ended comments, it seems that even though they had complaints, they found both interactive maps superior to paper maps—36 percent of them explicitly listed the maps' interactivity and responsiveness as main advantages. One of the Natlas users mentioned that he/she would skip the map service if not in an experimental situation and look for the answer to the question elsewhere. Indeed, it was very informative to see how several participants actually tried to do this when they became frustrated with the interface. We observed five Natlas viewers out of 15 (33 percent) and two Carto.net viewers out of 15 (13 percent), after getting frustrated with the map interfaces, who tried to leave the map web page and use other web search tools to find the answer to the test questions (e.g., Google, U.S. Census Bureau link, SVG source code).

One participant, responding to the Natlas interface, also offered a comment about interfaces in general: "Although it should not necessarily be the case always, I have a feeling that a poorly designed GUI (graphical user interface) is also an indication of a poorly implemented system." The following comment from another participant responding to the Carto.net interface user suggests: "They [legend information] did NOT always mean what I assumed. I think people don't like reading too much, and it would be best if the legends would match what most people assume by default." This comment ("I think people don't like reading too much") is particularly noteworthy as only one (7 percent) of our observed map users ever perused the Help button when they got stuck! A number of Carto.net viewers (40 percent) complained about the legends (legend design and legend description). This even included one participant claiming that the legend was wrong: "The legend and value reporting is wrongly designed."

Conclusions

We present an empirical-evaluation-based methodology that integrates eye movement analysis and traditional usability performance and satisfaction metrics for assessing interactive map interfaces. With a case study, we demonstrate that the information gathered from eye movement analysis can enhance usability studies both quantitatively and qualitatively. Overall results of SEE metrics and eye movement analysis for the case study confirm our hypothesis and indicate that users perform faster with Carto.net interface over Natlas. The study also reveals that the users, at least on two accounts, prefer Carto.net; more participants declared that they would like to use it frequently. However, the results obtained with the Natlas interface were more accurate, which indicates that both designs have usability problems, and the interfaces could both be improved. Usability researchers are well aware of the fact that professional deformation from years of training leaves the designer with little clue as to what is difficult for a nontrained person and what is easy-to-use.

The amount of data generated by modern eye-tracking devices is very high. Aside from the rapid progress in recent hardware and software technologies, simpler and more elegant solutions for processing large volumes of data are still needed. However, eye movement analysis provides us with information on visual behavior, which is commonly accepted as a proxy for mental attention (Webb and Renshaw 2008), and it is valuable for understanding how users make inferences with the interfaces.

In the demonstrated case study, eye movement data revealed microlevel usability issues regarding the *Identify* and *Redraw Map* buttons on Natlas as well as the mouse roll-over behavior on Carto.net, initially based on fixation durations, fixation counts, and differences from first fixation to first mouse clicks. Webb and Renshaw (2008) list a number of possible design recommendations that could be based on eye tracking (p.53). Within this frame, we believe we offered assistance by means of eye tracking to identify certain design problems (e.g., inefficient search with *Redraw Map* on Natlas, or mouse roll-over behavior on Carto.net, possible labeling problems with *Identify* button). We also believe we have observed and documented eye movements which may be highlighting participants' expectations (i.e., people spend time looking at the Menu area on Natlas when they are trying to locate a button to help them discover what *Identify* does).

Eye movement analysis is a complex endeavor for several reasons, including challenges that come with processing the large volumes of data and difficulties of interpreting eye movement behavior as it changes based on participant's state of mind or provided task and instructions. However, along with these challenges, it also provides insights that can enhance our understanding of how humans interact with machines, and in the scope of this paper, how humans interact with interactive map interfaces.

Future Work

This paper reports only part of the analysis. The issues left for follow-up include: individual task analysis, scan path analysis, and background training bias. We also plan to analyze participants who needed assistance separately from those who succeeded, and grouping participants based on their efficiency (studying patterns comparatively between low-performers and high-performers).

ACKNOWLEDGMENTS

We would like to offer our sincere thanks to our participants and the designers of Carto.net and Natlas interfaces for their time, patience, and kind cooperation, as well as to our proofreaders and reviewers for giving us valuable input to improve the quality and clarity of this paper.

REFERENCES

Bojko, A. 2006. Using eye Tracking to compare web page designs: A case study. *Journal of Usability Studies* 3(1): 112-20.

Brodersen, L. Andersen, H.H.K., and S. Weber. 2002. *Applying eye-movement tracking for the study of map perception and map design*. Publications Series 4, Volume 9. National Survey and Cadastre, Denmark.

Brooke, J. 1996. SUS: A "quick and dirty" usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., and A.L. McClelland (eds), *Usability Evaluation in Industry*, 1st ed. pp. 189-94. London, U.K.: Taylor and Francis.

Byrne, M.D. Anderson, J.R. Douglas, S., and M. Matessa. 1999. Eye tracking the visual search of clickdown menus. In: *Proceedings, ACM Human Factors in Computing Systems* (CHI '99), New York, New York. pp. 402-9.

Chin, J.P. Diehl, V.A., and K. Norman. 1988. Development of an instrument measuring user satisfaction of the human-computer interface. In: *Proceedings, ACM Human Factors in Computing Systems* (CHI '88), Washington, D.C. pp. 213-218.

Dix, A. Finlay, J. Abowd, G., and R. Beale. 2004. *Human-computer interaction*, 3rd ed. Harlow, Essex, England: Pearson Education Limited.

Duchowski, A.T. 2007. *Eye tracking methodology: Theory and practice*, 2nd ed. London, U.K.: Springer.

Fabrikant, S.I., S. Rebich-Hespanha, N. Andrienko, G. Andrienko, and D.R. Montello. 2008. Novel method to measure inference affordance in static small multiple displays representing dynamic processes. *The Cartographic Journal* 45(3): 201-15.

Fuhrmann, S., P. Ahonen-Rainio, R. M. Edsall, S. I. Fabrikant, E.L. Koua, C. Tobon, C. Ware, and S. Wilson. 2005. Making useful and useable geovisualisation: Design and evaluation issues. In: Dykes, M., A.M. MacEachren, and M.-J. Kraak (eds), *Exploring Geovisualisation*. Amsterdam, The Netherlands: Elsevier Ltd. pp. 553-66.

Goldberg, J.H., and X.P. Kotval. 1999. Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics* 24: 631-45.

Goldberg, J.H., M.J. Stimson, M. Lewenstein, N. Scott, and A.M. Wichansky. 2002. Eye tracking in web search tasks: Design implications. In: *Proceedings, Eye Tracking Research and Applications Symposium 2002*, New York, New York, USA. pp. 51-58.

Good, M., T.M. Spine, J. Whiteside, and P. George. 1986. User-derived impact analysis as a tool for usability engineering. In: *Proceedings, Human Factors in Computing Systems (CHI '86)*, NewYork, USA. pp. 241-46.

Haklay, M., and A. Zafiri. 2008. Usability engineering for GIS: Learning from a screenshot. *Special Issue on Use and User Issues The Cartographic Journal* 2(45): 87-97.

Harrower, M., and B. Sheesley. 2005. Designing better map interfaces: A framework for panning and zooming. *Transactions in GIS* 9(2): 77-89.

Hegarty, M., and D. Waller. 2004. A dissociation between mental rotation and perspective taking spatial abilities. *Intelligence* 32(2): 175-91.

Henderson, J.M., and F. Ferreira. 2004. Scene perception for psycholinguistics. In: Henderson, J.M., and F. Ferreira (eds), *The Interface of Vision, Language, and Action. Eye Movements and the Visual World*. New York, New York: Psychology Press. pp. 1-58.

Howard, D.L., and A.M. MacEachren. 1996. Interface design for geographic visualization: Tools for representing reliability. *Cartography and Geographic Information Science* 2(23): 59-77.

Irwin, E. 2004. Fixation location and fixation duration as indices of cognitive processing. In: Henderson, J.M., and F. Ferreira (eds), *The Interface of Vision, Language, and Action. Eye Movements and the Visual World*. New York, New York: Psychology Press. pp. 105-34.

Jacob, R., and K. Karn. 2003. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In: Hyönä, J. Radach, R., and H. Deubel (eds), *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*. Amsterdam, The Netherlands: Elsevier Ltd. pp. 573-605

Knapp, L. 1995. A task analysis approach to the visualization of geographic data. In: Nyerges, T.L., D.M. Mark, R. Laurini, and M.J. Egenhofer (eds), *Cognitive Aspects of Human-Computer Interaction for Geographic Information Systems*, 1st ed. Norwell, MA, USA: Kluwer Academic Publishers. pp 355-71.

Koua, E.L., A.M. MacEachren, and M.-J. Kraak. 2006. Evaluating the usability of visualisation methods in an exploratory geovisualization environment. *International Journal of Geographical Information Science* 20: 425-48.

Koua, E.L., and M.-J. Kraak. 2004. A usability framework for the design and evaluation of an exploratory geovisualisation environment. In: *Proceedings, The 8th International Conference on Information Visualisation (IV '04)*, London, U.K. pp. 153-58.

Larkin, J.H., and H.A. Simon. 1987. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science* 11(1): 65-100.

Lewis, J. 1995. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction* 7(1): 57-78.

MacEachren, A.M., and M.-J. Kraak. 2001. Research challenges in geovisualization. *Cartography and Geographic Information Science* 28(1): 3-12.

Medyckyj-Scott, D. 1993. Designing geographical information systems for use. In: Medyckyj-Scott, D., and H.M. Hearnshaw (eds), *Human Factors in Geographical Information Systems*, London, U.K.: Belhaven Press. pp. 87-100.

Nielsen, J. 1993. *Usability Engineering*. San Diego, California, USA: Academic Press.

Nivala, A.-M., S.A. Brewster, and L.T. Sarjakoski. 2008. Usability evaluation of web mapping sites. Special Issue on Use and User Issues of *The Cartographic Journal* 45(2): 130-40.

Nyerges, T.L. 1993. How do people use geographical information systems? In: Medyckyj-Scott, D., and H.M. Hearnshaw (eds), *Human Factors in Geographical Information Systems*. London, U.K.: Belhaven Press. pp 37-50.

Pretorius, M.C., A. P. Calitz, and D. van Greunen. 2005. The added value of eye tracking in the usability evaluation of a network management tool. In: *Proceedings, The 2005 Annual Research Conference of the South African institute of Computer Scientists and information Technologists on IT Research in Developing Countries (SAICSIT '05)*, White River, South Africa. pp. 1-10.

Robinson, A.C., J. Chen, E.J. Lengerich, H.G. Meyer, and A.M. MacEachren. 2005. Combining usability techniques to design geovisualization tools for epidemiology. *Cartography and Geographic Information Science* 32(4): 243-55.

Steinke, T.R. 1987. Eye movement studies in cartography and related fields. *Cartographica* 24(2): 40-73.

Tullis, T.S., and J.N. Stetson. 2004. A comparison of questionnaires for assessing website usability. In: *Proceedings, Usability Professionals Association Conference (UPA 2004)*, Minneapolis, Minnesotta. pp. 7-11.

Webb, N., and T. Renshaw. 2008. Eyetracking in HCI. In: Cairns, P., and A.L. Cox (eds), *Research Methods for Human Computer Interaction*. Cambridge: Cambridge University Press. pp. 35-69.

You, M., C.-W. Chen, H. Liu, and H. Lin. 2007. A usability evaluation of web map zoom and pan functions. *International Journal of Design* 1(1): 15-25.

TESTED WEB MAPS

Carto.net. 2008. An online interactive map interface showing the crime and poverty rates in the U.S.A. in 2000. [http://www.carto.net/papers/svg/us_crime_2000/; accessed 15/10/2008.]

Natlas. 2008. Map maker, an online interactive map interface of National Atlas of the United States. [http://nationalatlas.gov/natlas/Natlasstart.asp; accessed 15/10/2008.]

Cartography and Geographic Information Science (CaGIS):
Special Issue on Mapping Hazards and Disasters

SOLICITATION OF MANUSCRIPTS

In 1997, CaGIS published a special issue on "GIS and Environmental Risk Assessment" which primarily focused on topics related to environmental health. Since this time, interest in mapping hazards has expanded both conceptually and methodologically. Events such as the 2004 tsunami, Hurricane Katrina, the Sichuan earthquake, and the Southern California wildfires demonstrate the need for continued investigation of hazards and disasters, especially studies that create new knowledge of the multivariate spatial relationships that exist in these post-disaster environments. How can cartography, Geographic Information Systems (GIS), geo-technologies, and spatial analysis contribute to the generation of this knowledge? This special issue of CaGIS will target manuscripts that address two pervasive research gaps in mapping related to hazards and disasters:

1) Scale: Hazards and disasters are commonly studied at city or county scales, but research is showing that processes related to preparedness, response, recovery, and mitigation occur at a finer scale geography, such as the neighborhood. Papers that address approaches and data suitable for mapping vulnerability to events and recovery from these events at a local scale (anything finer than zip code) will be considered. Topics of interest include underserved populations such as elderly, children, disabled, new immigrants, migrant laborers, and those with chronic health conditions.

2) Timeframe: Most studies of extreme events have a limited time frame for monitoring long-term recovery (often 1 year). However, current non-spatial research suggests that the effects of exposure to these events have a longer timeframe of influence. Therefore articles will be targeted that address mapping and strategies for collecting temporally dynamic spatial data and those that focus on the display of space-time changes in the post-disaster environment. Additionally, little research exists on historical events and their application to current conditions; papers that take this approach are also of interest.

In addition to the two main foci of this special issue, papers that address approaches for dealing with the collection and display of disaster-related ephemeral and dynamic data, as well as mechanisms for disseminating the resulting spatial data to policy-makers and to the public will be considered.

In essence, the objective of this special issue is to highlight new knowledge gained through mapping hazards and disasters. All phases (planning, response, recovery, and mitigation) are requested, though studies that focus on long-term recovery and expansion of the concept of mitigation are of particular interest. Also, new forms of data collection, new types of spatial data, and approaches to display and dissemination are encouraged. The definition of hazards and disasters is broad, including geophysical, hydro-metereological, and human/technological issues, as well as health/medical events. Multi-disciplinary submissions are encouraged. Contributions from hazards and disasters both within and outside the USA are also welcomed. If you have any questions, please contact Dr. Jacqueline W. Mills at jacqueline.mills@usc.edu.

The special issue is scheduled for publication in January 2010. The timeline for submission follows:

1) Manuscripts of between 5,000 and 7,000 words should be submitted by May 24, 2009. Each paper will receive comments from three reviewers. Please send as a Word document to Dr. Jacqueline W. Mills at jacqueline.mills@usc.edu
2) Notification of acceptance will be sent via e-mail on August 23, 2009
3) Manuscript revisions will be completed by November 22, 2009
4) Publication is scheduled for January 2010

For information on manuscript format and style, please see the CaGIS guidelines for authors:
http://www.cartogis.org/publications/document.2006-09-05.7116381016