

Enhancing the Informatics Evaluation Toolkit with Remote Usability Testing

Brian E. Dixon, MPA

Regenstrief Institute, Indianapolis, IN

Indiana University, School of Informatics, Indianapolis, IN

Abstract

Developing functional clinical informatics products that are also usable remains a challenge. Despite evidence that usability testing should be incorporated into the lifecycle of health information technologies, rarely does this occur. Challenges include poor standards, a lack of knowledge around usability practices, and the expense involved in rigorous testing with a large number of users. Remote usability testing may be a solution for many of these challenges. Remotely testing an application can greatly enhance the number of users who can iteratively interact with a product, and it can reduce the costs associated with usability testing. A case study presents the experiences with remote usability testing when evaluating a Web site designed for health informatics knowledge dissemination. The lessons can inform others seeking to enhance their evaluation toolkits for clinical informatics products.

Introduction

Clinicians and other health care professionals interact with highly complex mix of clinical applications and Web sites in the provision of modern health care.¹ Patients too are interacting with a variety of Web sites and devices to manage their personal health. Many of these sites, applications, and devices are poorly designed.² Poor design can lead to workarounds and sometimes total abandonment. Such behavior in response to poorly designed health information products can be detrimental to health care quality, safety, and efficiency.

Human factors engineering techniques have been suggested as potential methods to assist developers improve the design of health informatics products.³ These techniques include testing (measurement and analysis of) a product's usability, or 'the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.'⁴

To be effective, usability testing should involve "informal" and "practical" solutions that do not require extensive training in usability for the average developer.^{5,6} Furthermore, usability testing should be routine and longitudinal, evaluating the design and user experience before, during, and after implementation.^{7,8}

Integrating usability testing into existing development and product lifecycles, however, is challenging. Cunliffe identified a number of well-known challenges, including a lack of accepted guidelines, a lack of knowledge about usability amongst developers, and few good models of user behavior.⁶ There are also challenges in the time and cost associated with the use of a formal usability lab and recruiting a meaningful number of users. Furthermore, products aimed at national or large scale deployment face challenges in testing with appropriately diverse samples drawn from a range of target populations.

In this paper, a case study involving usability testing done for a public sector, health informatics related Web site is presented. Limited resources and time prevented the use of traditional, laboratory-based usability testing as part of the site's ongoing evaluation strategy. An alternative form, remote usability testing, was employed. The methodology used is presented, and the lessons learned from the experience are discussed. The case study concludes with thoughts on wider adoption and use of remote usability testing as a practical component of routine processes to improve the design of health information products, including Web sites, applications, and devices. The objective of the paper is to describe the use of remote usability testing as a valid evaluation method. The author hopes that other informaticians will consider adding it to their evaluation toolkits and employ it to improve future development and use of clinical informatics products.

Background

The author was involved with the creation of a new Web site in the public sector, serving primarily as the site's information architect. The site was developed by a team composed of health informatics specialists and Web developers from academic and private industry. The author continues involvement in the evolution of the site as the ongoing information architect and day-to-day knowledge manager.

The Web site was designed to provide health care professionals, including physicians, nurses, practice managers, health information managers, quality managers, technology professionals, and public administrators with information resources related to the adoption, use, and implementation of health information technology (health IT). The Web site was to be a part of a larger campaign to help health care

provider organizations learn about various health information technologies and how the adoption and use of these technologies can improve health care quality, safety, efficiency, and effectiveness.

Traditional, laboratory-based usability methods were used during the formative development of the Web site in early 2006. A usability lab managed by the U.S. government and located in Washington, DC was chosen as the testing site. A usability consultant familiar with the lab was hired to facilitate the testing. Users were selected from several federal agencies and non-profit organizations in the DC area. Observers from the site's development team were present to watch the testing and take notes.

The development team collected several forms of feedback during the testing. Users were asked to complete a pre-test questionnaire on which they provided data about their Internet habits and prior knowledge of health information technologies. Users were then asked to complete three tasks randomly selected from a list of more than a dozen.

As they interacted with the Web site to complete the tasks, the facilitator probed users for their impression of the site, including their opinions about the depth of information available and the ease of information retrieval. This is often referred to as the "think-aloud" method of usability testing.⁹ The time to completion of tasks was measured, and the user's mouse-clicks and navigation were captured. When the user had completed the assigned tasks, then a post-test questionnaire was administered to measure the user's impression of the site, its functionality, and its overall design. Each test session was recorded for review and further analysis by the development team.

The usability testing provided a rich set of data and feedback the development team used to improve the site prior to launch. Certain sections of the site were re-organized, headings were modified, and additional icons were added to improve the design and architecture. Additional content was also created to provide more depth in response to user comments. The development team and its client both considered the usability testing to have been a success.

Once the site went live, the development team was tasked to create an ongoing, comprehensive evaluation strategy to monitor and improve the site over time. Like most Web teams, the initial strategy focused on collecting and reporting basic web analytics, such as the number of unique visitors, click-through statistics, and the number of transactions logged.¹⁰ The plan lacked any measurement of the user experience and few methods for collecting user feedback.

To improve the site's evaluation plan, the team turned to the recommendations of Wood et al.⁷ However, the

team found that a comprehensive strategy for measuring usage and the user experience can be costly and time consuming. Because of limited funding following the go-live and other resource constraints (a reduction in the number of development team members), the team sought to develop a strategy that provided "evaluation on a shoestring budget."¹¹

Methods

Remote usability testing is an established method that had been demonstrated to be comparable with traditional laboratory-based testing.^{12,13} In addition to being similar in nature to lab-based approaches, remote usability testing is often less costly when sampling users from a large geographic area.¹⁴

Remote usability testing comes in two primary forms: synchronous and asynchronous. In synchronous remote testing, the facilitator and user are separated spatially.¹⁵ They still interact with one another in real-time during the testing, via video (e.g., Webcam) or audio (e.g., headset, phone).

In asynchronous testing, the user performs tasks in the absence of a facilitator. Data are logged and recorded by the computer and sent to the testing team after the test. Andreasen et al. found a much richer body of evidence exists for synchronous remote testing versus asynchronous testing.¹⁶ Furthermore, fewer differences exist between synchronous remote testing and traditional testing when compared with asynchronous testing methods.

17 people were successfully recruited to participate in the synchronous remote testing using a protocol modified from the original, laboratory-based protocol developed a year prior. 12 men and five women each completed three tasks randomly chosen from a list of eight possible "top tasks."¹⁷ (Table 1)

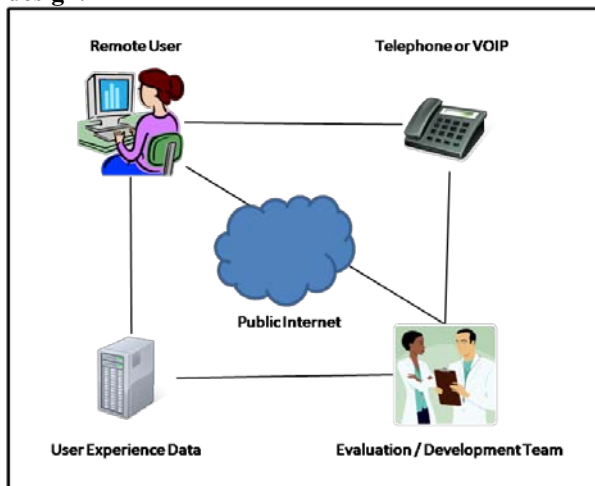
Task	Description
1	Locate information about a project in California trying to implement chronic disease registries.
2	Download a peer-reviewed journal article submitted from a project working on deploying a telehealth system.
3	Download a toolkit designed to help organizations evaluate outcomes from implemented health IT systems, such as electronic health record systems (EHRs).
4	What does the page on small and rural communities say is the biggest challenge when adoption health IT systems?
5	You are interested in implementing a Computerized Provider Order Entry (CPOE) system. You're specifically interested in the challenges others have had in using this type of

	system. What do some CPOE projects say is a major challenge involved in alerting clinicians to potential errors when entering orders using a CPOE system?
6	Locate information on physician use of the PDA and other handheld devices. What kinds of tasks do physicians most often use the devices to accomplish?
7	Locate the due date for a new funding opportunity in the area of health IT.
8	List the major U.S. initiatives trying to help providers get connected to share data.

Table 1. The top tasks from which users were randomly assigned three (3).

Remote users performed assigned tasks through the WebEx Communications (Santa Clara, CA) platform over the public Internet under supervision of a facilitator. The shared Web browser functionality of WebEx enabled the facilitator to synchronously view the remote user's browser, seeing what the remote user saw. In addition, the WebEx functionality allowed the facilitator to observe the remote user's mouse movements as he or she navigated the site in performance of a given task. The facilitator and user communicated using a standard telephone or voice-over-IP (VOIP), depending on the user's location and preferences. Each session was recorded, and the user's responses to the pre-test and post-test questionnaires were captured by WebEx and posted to a research server at the Regenstrief Institute. (Figure 1)

Figure 1: Synchronous remote usability testing design.



The remote usability session was initialized by the facilitator. After joining the session, the facilitator began the testing by asking the participant to answer several pre-test questions. Once introductions and pre-test information had been completed, the facilitator

opened a shared Web browser and navigated to the home page of the site under examination. The user then took over navigation of the site, and the facilitator asked the user to complete three randomly selected tasks for the session. The user was encouraged to "think aloud" as he or she completed the tasks. The facilitator also probed for further information when the user appeared to be stuck or contemplating his or her next action. When the three tasks were completed, the facilitator asked the user to complete a post-test questionnaire. An open-ended post-test interview was also performed where the user had an opportunity to verbally express comments about the testing experience.

Quantitative and qualitative data were collected during each session. The primary instrument for measuring the user experience was the system usability scale (SUS).¹⁸ The SUS is a widely used and scientifically validated usability instrument. The SUS produces a usability score between 0-100, based on respondents' component scores.

Qualitative data collected included notes from the facilitator, the perceptions of the user during the testing, and thoughts of the participants after they had completed the testing. Pre and post-test questionnaires also provided information on the user's Internet usage, knowledge of health IT, and other relevant demographic details.

Results

The results from the remote usability testing provided the development team with data on how well the site supported a positive user experience.

The SUS values in the remote usability tests ranged from 5 - 87.5 with a standard deviation of 27.27, indicating a high degree of variability in respondents' answers. Excluding outliers, the instrument had a mean SUS score (N=12) of 74.38 – a "very good" score – and a standard deviation of 11.08 – much lower variability. The outliers were participants who primarily used the site's search functionality as opposed to browsing for content using the site's navigational components. This distinction revealed to the development team that the site's search component needed significant improvement.

In addition to the SUS, the user experience data included qualitative responses provided during the performance of the top tasks. This data revealed several patterns. First, many users failed to recognize the site as being focused on health informatics topics. Instead, they associated the site with the government agency funding the site and perceived the site to be that agency's main Web site. Second, users perceived many of the site's pages as too densely populated with

text. The users asked for more images and more pathways that would allow a quick scan of information followed by a click to access additional details. This feedback provided the development team with direction for enhancing the site's usability in the next budget cycle.

The average participant (N=17) was between 31-40 years of age, held a Bachelor's degree, served as either a clinical or IT manager, and considered themselves knowledgeable about the Web and health IT issues. When asked about the Web sites they used on a regular basis, nine participants indicated they used Google. The next most popular Web site was the organizational site for the Health Information Management Systems Society (HIMSS). Other Web sites mentioned by users included professional societies such as the American Medical Association (AMA), media companies such as CNet and CNN, and health IT vendors such as Eclipsys and Cerner. Most users received their health IT news from trade publications such as Healthcare IT News and Modern Healthcare.

Discussion

The results from the remote testing methods were generally comparable to those of the original, laboratory-based testing. Comparing the two sets of SUS scores revealed the overall usability of the Web site had increased, which was expected since adjustments had been made and refined over the site's first year in operation. The development team also received useful feedback regarding future work to improve the site's usability in the next budget cycle. The team concluded from the experience that remote testing was just as helpful as face-to-face testing in a laboratory environment, a conclusion that supports previous studies comparing various usability methodologies.^{13,19}

Beyond its comparative effectiveness, the team found that remote usability testing was cost-effective. The team was able to conduct twice as many tests at a lower cost. Fewer resources needed for usability testing enabled the team to conduct additional analyses of the site, including face-to-face focus groups, with no increase in the total budget used for site evaluation. This is important to public and small organizations, such as federally qualified health centers and community hospitals, which often have very small budgets for ongoing assessment and evaluation activities.

Remote testing further enabled the collection of data from a wide variety of locations and target audiences. Recall that the site aims to provide information resources to health care professionals across the nation working in a variety of organization types. Although the participants were in no way representative of the

U.S. health care system or any one component of it, remote usability testing enabled the development team to reach a more diverse group than its original testing. Remote testing holds great promise for the recruitment of more diverse user populations when developing or enhancing an application or Web site.

Finally, remote testing allowed the examination of user experiences in a real-world environment. Most users logged into the testing environment from their primary office or home computers, natural settings where these users would normally interact with the Web site under evaluation.⁸ Screen resolutions, browsers, keyboards, mice, etc were all tested as they are in users' daily settings. This allows usability and informatics professionals to evaluate products with a broader range of configurations than those available in the laboratory environment. It further makes users more comfortable, and it can illuminate issues germane to a user's environment.¹²

Given its comparable performance with traditional, laboratory-based usability testing, development teams and evaluators should consider adding remote usability testing to their evaluation toolkits for clinical informatics products. In general, usability testing and other human factors techniques are recommended for use during the development of products used in health care environments given the potential impact on quality of care and patient safety.¹⁻³ Furthermore, these techniques should be used post-implementation, guiding the evolution of a product as it matures.⁷

Based on the results of the case study presented here, remote usability testing has the potential to be an effective tool for measuring the user experience throughout a product's development and evolution. The method may support additional interactions with a greater number of users in more natural settings using the same amount of resources (potentially less) than other usability testing techniques. Use of this method in other informatics products is recommended to validate this assumption.

Although appropriate in many situations, remote testing may not always be the ideal usability testing method. A carefully controlled research experiment, for example, may prohibit the use of remote testing because researchers might not have control over the environment in which the user interacts with the interface. Similarly, interfaces designed for use side-by-side with other applications might not be suitable for remote usability testing strategies.

Conclusion

Ongoing evaluation is critical to the long-term success of a clinical informatics product. Many development teams struggle, especially in the public and non-profit

sectors, to conduct ongoing evaluation of the user experience due to small budgets, few personnel, or lack of experience with usability testing. When budgets, time, and knowledge are constrained, many organizations employ only basic usage measures and analytics. Ignoring the user experience leads to poorly designed products, this may negatively impact quality and safety. It further prevents evaluators from painting a comprehensive picture of usage and adoption, preventing measurement of the potential impacts on quality and safety.

While important steps have been taken to develop and promote remote testing methods, additional work is encouraged to make remote testing more integrated into routine usability testing strategies for health informatics products. Education, research, and development of a wide range of integrated remote testing solutions is necessary to share remote testing methods with usability and informatics professionals, evaluate the application of these methods, and enable organizations to quickly deploy remote testing techniques in the real-world.

Acknowledgements

This paper is derived from work supported under a contract with the Agency for Healthcare Research and Quality (290-04-0016). The author would like to thank Jason Saleem, PhD, for his comments on a draft of this paper. The author would also like to thank Atif Zafar, MD, Julie McGowan, PhD, and Cait Cusack, MD, for their support and encouragement.

References

1. Karsh BT. Beyond usability: designing effective technology implementation systems to promote patient safety. *Qual Saf Health Care*. 2004 Oct;13(5):388-94.
2. Johnson CW. Why did that happen? Exploring the proliferation of barely usable software in healthcare systems. *Qual Saf Health Care*. 2006 Dec;15 Suppl 1:76-81.
3. Carayon P. Human factors of complex sociotechnical systems. *Appl Ergon*. 2006 Jul;37(4):525-35.
4. Bolchini D, Finkelstein A, Perrone V, Nagl S. Better bioinformatics through usability analysis. *Bioinformatics*. 2009 Feb 1;25(3):406-12.
5. Nielsen J. User interface directions for the Web. *Communications of the ACM*. 1999;42(1): 65-72.
6. Cunliffe D. Developing usable Web sites – a review and model. *Internet Research: Electronic Networking Applications and Policy*. 2000;10(4):295-307.
7. Wood FB, Siegel ER, LaCroix EM, Lyon BJ, Benson DA, Cid V, Fariss S. A practical approach to e-government Web evaluation. *IT Professional*. 2003 May-Jun;5(3):22-28.
8. Cuddihy E, Wei C, Barrick J, Maust B, Bartell AL, Spyridakis JH. Methods for assessing Web design through the Internet. In: Van der Veer G, Gale C, editors. *CHI '05 extended abstracts on Human factors in computing systems*; 2005 Apr 2-7; Portland, Oregon. ACM Press; 2005. p. 1316-1319.
9. Jaspers MWM, Steen T, van den Bos C, Geenen M. The think aloud method: a guide to user interface design. *IJMI* 2004;73(11-12):781-795.
10. Phippen A, Sheppard L, Furness S. A practical evaluation of Web analytics. *Internet Research*. 2004;14(4):284-293.
11. Poon E. Evaluation on a shoe string [Presentation]. AHRQ National Resource Center for Health IT. 2007.
12. West R, Lehman KR. Automated summative usability studies: An empirical evaluation. In: Grinter R, Rodden T, Aoki P, Cutrelle, Jeffries R, editors. *Proceedings of the SIGCHI conference on Human Factors in computing systems*; 2006 Apr 22-27; Montreal, Quebec, Canada. ACM Press; 2006. p. 631-639.
13. Tullis T, Fleishman S, McNulty M, Cianchette C, Bergel M. An empirical comparison of lab and remote usability testing of Web sites. In: *Proceedings of the Usability Professionals' Association Annual Conference*; 2002 Jul 8-12; Orlando, Florida. p. 32.
14. Dray S, Siegel D. Remote possibilities? International usability testing at a distance. *Interactions*. 2004 Mar-Apr;11(2):10-17.
15. Brush AJ, Ames M, Davis J. A comparison of synchronous remote and local usability studies for an expert interface. In: Dykstra-Erickson E, Tscheligi M, editors. *CHI '04 extended abstracts on Human factors in computing systems*; 2004 Apr 24-29; Vienna, Austria. ACM Press; 2004. p. 1179-1182.
16. Andreasen MS, Nielsen HV, Schroder SO, Stage J. What happened to remote usability testing? An empirical study of three methods. In: *Proceedings of ACM CHI 2007 Conference on Human Factors in Computing Systems*; 2007 Apr 28 – May 3; San Jose, California. ACM Press; 2007. p. 1405-1414.
17. Koyani S. Cross-governmental perspective on usability. In: *Proceedings of the Usability Professionals' Association Annual Conference*; 2008 Jun 16-20; Baltimore, MD.
18. Brooke J. SUS – A quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. *Usability evaluation in industry*. Bristol, PA: Taylor and Francis Ltd; 2003. p. 189-192.
19. Thompson KE, Rozanski EP, Haake AR. Here, there, anywhere: remote usability testing that works. In: Helps R, Lawson E, editors. *Proceedings of the 5th conference on Information technology education*; 2004 Oct 28-30; Salt Lake City, Utah. ACM Press; 2004. p. 132-137.