

Usability@90mph: Presenting and Evaluating a New, High-Speed Method for Demonstrating User Testing in Front of an Audience

Paul F. Marty
College of Information
Florida State University

Michael B. Twidale
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

Abstract

This article documents the authors' attempt to develop a quick, inexpensive, and reliable method for demonstrating user testing to an audience. The resulting method, *Usability@90mph*, is simple enough to be conducted at minimal expense, fast enough to be completed in only thirty minutes, comprehensible enough to be presented to audiences numbering in the hundreds, and yet sophisticated enough to produce relevant design recommendations, thereby illustrating for the audience the potential value of user testing in general. In this article, the authors present their user testing demonstration method in detail, analyze results from forty-four trials of the method in practice, and discuss lessons learned for demonstrating user testing in front of an audience.

Keywords

User Testing Demonstrations; Usability Analysis; High-Speed, Low-Cost User Testing.

Introduction

Under the right conditions, only a small moment of time is necessary to demonstrate a lesson well known to usability professionals, that the best way to evaluate an interface for usability is to test that interface with representative users. Watching users struggle with an unfamiliar interface, even observers who have never considered the importance of usability become instantly aware of the need for and benefits of representative user testing (Dumas, 2002; Nielsen, 1994a). As researchers and practitioners call for increased accountability from designers in terms of meeting the needs of all users (Shneiderman, 2000), it is crucial that individuals from every discipline become aware of the value of user testing for improving the usability of information interfaces. There is no doubt that user testing demonstrations can be an extremely powerful way of illustrating the potential benefits of usability analysis to a wide variety of observers.

The problem, however, is that demonstrating user testing in front of an audience is a non-trivial procedure. There are two major difficulties with conducting user testing demonstrations. First, neither experimental, laboratory-based methods of user testing nor simulations of user testing suitable for college classrooms lend themselves well to audience demonstrations. While

classroom simulations can be performed quickly and cheaply, it is rare that simulations produce relevant results or useful recommendations for design improvements on their own, thereby limiting the potential impact of this approach on an audience. Similarly, laboratory-based user testing, despite its high potential as a demonstration tool, tends to be both expensive and time-consuming, and the typical usability laboratory is not a suitable environment for demonstrating user testing to an audience greater than a few people.

The second difficulty with demonstrating user testing to an audience is that observing users in action is only part of the process; if people are truly to understand the benefits of user testing, they need to acquire at least a basic understanding of the entire process from start to finish. The complete user testing process requires trained usability evaluators to study an interface (often one with which they are unfamiliar), assess its strengths and weaknesses, develop representative scenarios of use, administer these scenarios to representative users, analyze and evaluate the results, and generate relevant and useful recommendations for design improvements (Nielsen, 1994b; Rubin, 1994; Hackos and Redish, 1998). Usability researchers and practitioners have found that this process can take days if not weeks to complete, far too long a time for a demonstration (Thomas, 1996). For this reason, usability professionals wishing to prove the value of user testing to their colleagues have traditionally had to rely on “guerrilla” tactics that gradually encourage usability engineering practices over long periods of time (Nielsen, 1994a). Given these constraints, a relatively quick, inexpensive, and easy-to-use method of demonstrating user testing to an audience would be a valuable tool for individuals wishing to encourage a “culture of usability” in their own organizations (cf. Fraser, 2002).

This article, therefore, addresses the following research question: Can a high-speed user testing demonstration method be developed that will help audience members understand the value of user testing quickly, yet without sacrificing the inherent “realism” of user testing by relying solely on simulations? To answer this question, the authors developed a new method capable of demonstrating an entire user test from start to finish in thirty minutes, and tested this method by evaluating websites in front of large audiences at several different national and international conferences. Developing this method was challenging, as it had to be general enough to reach audience members with varying levels of technical expertise from almost any discipline, yet specific enough to generate relevant and useful (i.e. not simulated) design recommendations that could potentially improve the usability of the interfaces being tested.

The resulting method, which we call *Usability@90mph*, is simple enough to be conducted at minimal expense, fast enough to be completed in only thirty minutes, comprehensible enough to be presented to audiences numbering in the hundreds, and yet sophisticated enough to produce relevant and useful design recommendations, thereby illustrating for an audience the potential value of user testing in general. Over the past four years, we have tested this method forty-four (44) times by demonstrating it to audiences at six different national and international conferences for museum professionals. Each demonstration was conducted on a different museum website, and each test resulted in design recommendations that, if implemented, would likely improve the usability of those websites (for examples, see Marty and Twidale, 2004). Of greater significance for this article, however, is that the results of our tests clearly indicate that this demonstration method has the potential to introduce diverse audiences to the power, value, and benefits of user testing, quickly, cheaply, and reliably.

In this article, we will review the case for usability demonstration methods, present the *Usability@90mph* method in detail, analyze the results from our forty-four tests of the method, and discuss the lessons learned for demonstrating user testing in front of an audience.

Demos@90mph: The Case for High-Speed User Testing Demonstrations

The past ten years have seen great advances in the willingness of most organizations to concede the value of usability engineering for improving their products (Dumas, 2002). The overall usability of websites, for example, continues to improve as a direct result of more attention being paid to user testing by design companies (Nielsen, 1999). Nevertheless, misconceptions about the value of user testing persist, and consumers still contend daily with poorly designed and unusable interfaces (Shneiderman, 2002). Even today, many need to be convinced of the value of user testing for improving interfaces (Donahue, 2001; Siegel, 2003).

Fortunately for advocates of usability analysis, user testing is itself already a very powerful way of convincing people of the benefits of usability engineering. Often, the mere act of watching a user test in process is all that is necessary to convert individuals skeptical of the benefits of user testing. As Dumas (2002) writes,

“One of the important assets of testing is that it sells itself. Watching a few minutes of live testing can be very persuasive. ... When people see their first live test they are almost always fascinated by what they see. They gain an understanding of the value of the method. ... Some of them will even become advocates for testing.” (p. 1103)

Nielsen (1994a) has made similar arguments, while also pointing out the potential difficulties of employing this approach in organizations:

“Many teachers of usability engineering have described the almost religious effect it seems to have the first time students try running a user test and see with their own eyes the difficulties perfectly normal people can have using supposedly ‘easy’ software. Unfortunately, organizations are more difficult to convert, so they mostly have to be conquered from within by the use of guerrilla methods like discount usability engineering that gradually show more and more people that usability methods work and improve products.” (p. 269)

Many of the “discount usability engineering” techniques employed to demonstrate the value of usability analysis are “inspection methods,” such as heuristic evaluation or cognitive walkthroughs, where interfaces can be evaluated without the involvement of actual users (Nielsen and Mack, 1994). While such techniques are extremely valuable for usability analysts, they are not typically as dramatic (from an observer’s perspective) as representative user testing can be. If one had an eye-opening method of easily and effectively demonstrating user testing to large audiences, one could potentially convert more people to the benefits of user testing more quickly; and the faster such a method worked, the more easily demonstrators could maximize audience impact per time expended.

Developing such a method presents certain challenges. The need to demonstrate user testing in front of a large audience (on stage using large screen computer projections, for instance) places serious constraints on the method. Taking the user test out of the user testing lab can be problematic for many evaluators, especially since experimental user testing methods often rely on non-portable, proprietary hardware or particular physical surroundings (Rowley, 1994; Wixon and Ramey, 1996). Moreover, given the public nature of the method, it might prove difficult for evaluators to get permission to use video cameras or other recording devices to document the tests, thereby limiting the data analysis capabilities.

The most serious challenge, however, is the challenge of time. Every stage in user testing can be time consuming. Nielsen (1994a) estimates that conducting one user test, including evaluating the interface, planning the test, coming up with representative tasks, administering those tasks, evaluating the results, and making design recommendations, could consume at least 50 hours; for many usability engineers, conducting user tests quickly means limiting the testing process to three days (Bauersfield and Halgren, 1996). Given all of these challenges, is it possible to develop a user testing demonstration method that can be performed quickly while still illustrating the inherent value of user testing?

Research@90mph: Assessing a Method for High-Speed User Testing Demonstrations

Our pedagogic aims in answering this research question were to make the process of user testing visible to and comprehensible for as large an audience as possible. It was critical to reach out not just to those already interested in user testing and already practicing it in their organizations, but also those who wanted to do so but were not sure how to do it under extreme constraints of time and budget; those who wanted to advocate for user testing in their organizations and needed help doing that; those who had heard about user testing but wanted to know a bit more; those who were skeptical that user testing could be performed out of a laboratory environment; as well as those who believed that user testing was too expensive and beyond the budget of their organization. As a result we needed to create a somewhat light, somewhat entertaining method that could convey the concepts of user testing while simultaneously providing suggestions for ways in which user testing can be conducted at minimal cost. Nevertheless, our principal goal was to encourage further consideration of user testing amongst audience members, rather than to present a pre-packaged solution for user testing which others could implement in their own organizations.

As instructors who regularly teach courses in usability analysis, we were well aware of the power of usability demonstrations when we began this project; we were also inspired by the *Interactionary* performed at CHI conferences as an entertaining way of illustrating design concepts in very short time frames (Berkun, 2003). Our first step in developing our method for user testing demonstrations, therefore, was to isolate the key principles of representative user testing (system analysis, task analysis, user testing, results analysis, design recommendations) and distill them to the barest possible elements that could be demonstrated in the quickest possible time. The result, which we called *Usability@90mph*, was a simple, step-by-step process where evaluators spend ten minutes assessing a previously unknown interface and developing representative tasks, ten minutes administering these tasks to representative users, and ten minutes analyzing the results of the tests to identify usability flaws and make recommendations for design. Each ten minute section was further broken down with specific rules that governed

the actions of the evaluators and all participants in the method (see below for a detailed description of the operation of the method).

Our next step was to evaluate the usefulness of this method as a proof of concept for demonstrating the user testing process to an audience. We were especially curious to see whether a single demonstration, lasting thirty minutes from start to finish, could generate design recommendations of sufficient number to illustrate the potential of user testing. To be a genuine demonstration of the power of user testing, *Usability@90mph* could not be merely a *simulation* of user testing (generating stock or pre-determined answers), but had to be a quasi-legitimate form of user testing, producing specific, individual design recommendations for each interface tested. Although we could never claim in thirty minutes to have found the majority or even the most serious of the usability flaws associated with any given interface, the flaws that we did find had to be accepted by the audience as legitimate usability problems with the interface being evaluated; only then would the method serve its purpose as a demonstration tool.

To assess the capabilities of *Usability@90mph*, we conducted 44 separate trials of the method at six different national and international conferences for museum professionals over the past four years. The total amount of time dedicated to the method at each conference varied, ranging from a minimum of two hours to a maximum of six hours. In sum, a total of twenty-two hours were allocated to the demonstration and testing of the method, and forty-four high-speed user testing demonstrations were conducted in this amount of time. The size of the audiences varied widely from conference to conference, from a minimum of about thirty to a maximum of around two hundred. Although an exact count was impossible to determine, it is safe to say that over the course of the 44 trials, at least six hundred people observed at least one user testing demonstration.

The fact that we presented and assessed our method at museum conferences for museum professionals using museum websites was a somewhat arbitrary choice based on the authors' research interests. The websites that were evaluated with this user testing demonstration method were suggested by conference attendees (typically users of museum websites, developers of museum websites, or both) through the use of a signup sheet; a representative from the organization whose website was being evaluated was required to be on hand during the user testing demonstration. Volunteer user testers were selected from the audience at the start of each session, provided they had no prior familiarity with the specific museum website being evaluated. In order to learn what the volunteer user testers were thinking, we employed two common techniques (usually alternating them from test to test). In 24 of 44 trials, we used the "think-aloud" protocol where one volunteer user vocalized his or her thoughts during the test (Waes, 2000). In 20 of 44 trials, we used a variation on "constructive interaction" where two volunteers discussed strategies of completing tasks in front of the audience (Wildman, 1995).

Conducting research at such high speeds is not easy, and the public nature of these events combined with the impossibility of obtaining informed consent from all audience members attending these conferences meant that it was impossible for us to create audio or video recordings of the tests. We therefore had to rely on notes that allowed us to reconstruct each test at the end of each conference; while necessarily brief, these provided valuable data that helped us assess the method. During each session, we recorded the types of sites evaluated, the tasks

developed for each test, the number of tasks actually administered during the test, the number of tasks completed by the user testers, the usability flaws uncovered in each test, and design recommendations made by the evaluators. In addition, we gathered informal feedback from the participants in each demonstration by asking the site representatives, the volunteer user testers, and the audience whether they thought that the results of each evaluation represented valid usability problems that typical users might face. This feedback about the relevance of the usability flaws uncovered by the *Usability@90mph* method helped validate the potential of the method as a demonstration tool.

Using these data, we conducted quantitative and qualitative analyses to determine 1) the value of our user testing demonstration method in terms of how much we were able to accomplish in each thirty minute test; and 2) lessons learned about convincing audiences of the benefits of user testing through demonstration methods. We will now present our method, discuss the success of the method as a demonstration tool, and present the lessons we learned for making demonstrations of user testing as successful and as valuable as possible.

***Usability@90mph*: Presenting a Method for High-Speed User Testing Demonstrations**

Each *Usability@90mph* session is designed to take thirty minutes, although some give and take in this duration is expected. As mentioned above, the goal is to spend ten minutes analyzing the site to be evaluated and determining representative tasks, ten minutes running the tasks with representative users, and ten minutes analyzing the results with the help of the audience. At a minimum, the method requires two evaluators, one representative from the organization that owns the website being tested, one or two user testers, and an audience. The four distinct roles played by the participants are defined here:

Evaluators refers to the trained usability experts who run the user-testing demonstration. There are two evaluators: the macro-level evaluator and the micro-level evaluator; the former is responsible for developing and administering representative tasks while the latter is responsible for guiding the volunteer user testers through the user testing process.

Site Representative refers to the person representing the organization responsible for the site being evaluated. Site representatives need to be present when their sites are evaluated.

Audience refers to the observers of the user testing demonstrations, who also serve as potential volunteer user testers.

Volunteer User Testers refers to the individuals who have volunteered to perform a representative task or scenario which a typical visitor to the site might naturally attempt to accomplish; there can be one or two volunteer user testers in each evaluation.

Usability@90mph requires only a minimal amount of preparation, and assumes that the demonstration area has been prepared in advance with a computer and a projector. The websites to be evaluated should also have been determined ahead of time, and it is preferable that these websites be unfamiliar to the usability evaluators, so as not to compromise the integrity of the demonstration. The successful implementation of the *Usability@90mph* method involves the

completion of three distinct stages: task analysis and development, testing of site with tasks, and analysis of test results. The following discussion presents the actions of the participants throughout these three stages (see Table 1 for a summary of this discussion).

Time (min)	Method Action Item	Site Representative	Evaluator (Macro)	Evaluator (Micro)	Volunteer Tester(s)	Audience
<i>Stage 1: Analysis of Site and Task Development (10 minutes)</i>						
2	Introduction	Moves to front of room	Projects site on screen; asks for volunteers from audience.		Leave room w/ evaluator.	Volunteers; Observes
4	Site Presentation	Introduces Site to Audience	Observes and plans tasks	Outside of room, micro-level evaluator instructs volunteer user testers about their roles and responsibilities in the user testing process.		Observes
4	Task Development	Macro-level evaluator develops representative tasks that meet with approval of site rep.				Observes
<i>Stage 2: Administration of Tasks to User Tester(s) (10 minutes)</i>						
10	Testing of Site with Tasks by Users	Observes	Administers tasks to user testers	Guides volunteer user testers during test	Perform representative tasks on site	Observes
<i>Stage 3: Analysis of Results, Usability Flaws, and Design Recommendations (10 minutes)</i>						
10	Analysis of Test Results	Discusses value of suggested improvements	Discusses observed usability flaws and presents design recommendations		Answer questions; Suggest improvements	Suggests Improvements

Table 1: Summary of *Usability@90mph* Method

Stage 1: Task Analysis and Development (10 minutes)

Assuming all participants are in attendance and ready to proceed, the evaluators can begin the first stage of this demonstration method with virtually no preliminary effort. This stage has three main steps: introduction, site presentation, and task development.

Introduction (2 minutes)

The evaluators introduce *Usability@90mph* by telling audience members that the purpose of this demonstration is to illustrate how observing users attempting to perform representative tasks with a given website can highlight hitherto unknown usability problems with that site. The site to be evaluated is announced and the site representative comes forward to the front of the room. The evaluators then ask the audience for volunteer user tester(s), with the stipulation that those volunteering can have no prior experience with the website being evaluated. One evaluator (the micro-level evaluator) then takes the volunteer user tester(s) out of the room. The site to be evaluated is then projected on the screen for the audience.

Site Presentation (4 minutes)

Once the micro-level evaluator and volunteer user tester(s) have left the room, the other evaluator (the macro-level evaluator) asks the site representative to quickly introduce this site to the audience. Depending on the preferences of the site representative, the evaluator, the site representative, an audience volunteer, or some other assistant can navigate the website on the computer in real time while the site representative describes the site to the audience. The evaluator should ask the site representative to concentrate on aspects of the site that would be of interest to the website's typical visitors. This step should take no more than four minutes.

Task Development (4 minutes)

After the site representative has introduced the site, the evaluator, drawing upon the site representative's presentation, develops several representative scenarios of use. These scenarios are phrased as tasks, and should reflect typical tasks that a representative user of this website would actually be interested in doing. These tasks can duplicate aspects of the site explicitly demonstrated by the site representative, or can be hypothetical tasks suggested by the evaluator. Suggestions for tasks from the audience may also be entertained. Once a list of tasks has been developed, each task should be presented to the audience and the site representative, and the site representative asked to confirm that each accurately represents tasks a typical visitor to this website might attempt. Tasks should be abandoned or modified, and new tasks developed, until the site representative is satisfied that the tasks are in fact representative. At this point in time, it may not be obvious to anyone except the site representative exactly how these tasks would (or would not) be completed by any given visitor to the site. The evaluator should aim to develop four or five representative tasks (approximately one task per minute).

While the task analysis is taking place in the room, the micro-level evaluator outside the room explains to the volunteer user tester(s) what will happen when they re-enter the room. He or she needs to explain to the volunteers that they are conducting a test on the website, and that they themselves are not being evaluated (a useful metaphor for this purpose is to explain that this is a "test drive" and not a "driving test"). It is especially important for the evaluator to make sure the volunteers feel at ease with the process of testing the website. If there is only one volunteer user tester, the evaluator will ask the volunteer to explain to the audience what he or she is thinking when he or she returns to the room to navigate the website. If there are two volunteers, the evaluator asks the two volunteers to discuss their thoughts and proposed actions when they return to the room to navigate the website.

Stage 2: Testing of Site with Tasks (10 minutes)

Once the representative tasks have been developed, the micro-level evaluator and volunteer(s) waiting outside the room are asked to return inside and to take a seat at the test computer. The macro-level evaluator stands where he or she can see the projected screen. It is at this point where the need for two evaluators becomes clear, as each assumes a different role in the user testing demonstration process.

The evaluator who developed the tasks, the macro-level evaluator, administers each task one at a time in the form of scenarios. Each scenario is presented to the volunteer(s) who then proceed to

use the website to complete the assigned task, making sure to vocalize their thoughts as the evaluators and audience members observe. The macro-level evaluator, who along with the audience has some familiarity with the website, follows the volunteer(s) interactions with the interface looking for macro-level usability flaws: unclear labeling, confusing site layout, poorly designed information architecture, difficult navigation, etc.

The evaluator who left the room with the volunteer(s), the micro-level evaluator, who like the volunteer(s) is seeing this website for the first time, observes the proceedings looking for micro-level usability flaws: confusions, hesitations resulting from unclear vocabulary, evidence of user surprise, mismatches between system terminology and user expectations, unclear or unexpected system feedback, mis-clicks with the mouse, verbal references, mouse hovering, etc. The micro-level evaluator is also responsible for prompting the volunteer(s) to “think-aloud” or “interact” if necessary.

Each evaluator takes notes about task completion rates, usability flaws encountered, and potential design recommendations throughout the test. The macro-level evaluator is responsible for determining whether a task has been successfully completed or whether it is likely the volunteer test users will not complete a given task and a new task should be assigned. The macro-level evaluator may choose to vary the order of the tasks depending on the results of the user test. The macro-level evaluator continues assigning tasks until all tasks have been completed or ten minutes have passed.

Stage 3: Analysis of Test Results (10 minutes)

At this stage, the volunteer(s) are thanked for their participation, and the two evaluators discuss the usability flaws they observed. Each evaluator will have a unique perspective on the problems with the website. How the evaluators present their findings is left to the discretion of the evaluators, although we found it useful for the micro-level evaluator to present his or her findings first. This allows for the evaluator without insider knowledge to summarize the nature of the problems a user of this website might experience. The macro-level evaluator can then compare the actions of the volunteer test users with the site representative’s expectations, and discuss the assumptions inherent to the design process.

At various points in this discussion, the evaluators may need to ask the volunteer test users to try to explain certain actions or choices they made during the user testing demonstration, carefully phrasing the request so that it is acceptable for them to venture a guess or say they do not remember. In addition, the evaluators should encourage audience members to contribute suggestions about usability problems they observed during the user test. The goal of this stage is to generate as many design recommendations as possible. These recommendations should directly address usability flaws discovered during the user test, and should be positive, constructive comments that will result in usability improvements of the website being tested. We found that once a usability problem has been identified and its potential causes clearly articulated, it is relatively easy for the evaluators, the volunteers, the audience, or the site representative to generate ideas that could fix the problem.

Once this is finished, the *Usability@90mph* method is complete. The evaluators should thank the site representative and the volunteer user tester(s) for participating in the demonstration. They are now ready to move on to their next thirty-minute, high-speed user testing demonstration.

Results@90mph: Analyzing the Success of High-Speed User Testing Demonstrations

As discussed above, our goal in this study was to develop and evaluate a high-speed method of demonstrating the user testing process that would help audience members better understand the value of conducting user tests in general. To be successful, the *Usability@90mph* method needed to show high numbers of tasks administered and design recommendations made per unit time as well as by the number of audience members reached during the demonstration. Our data analysis clearly illustrates how much of the user testing process can be demonstrated in a small amount of time to a large number of people.

In only twenty-two hours worth of demonstrations, working with unfamiliar interfaces in front of audiences totaling over 600 people, we developed a total of 178 representative tasks and administered 141 of them to representative users, who were then able to complete 100 of those tasks during the tests. Moreover, the results of these tests enabled us to make more than 500 design recommendations, each directly related to at least one usability flaw found while the volunteers were attempting to complete the administered tasks (see Table 2).

<i>Tasks Developed</i>	<i>Tasks Administered</i>	<i>Tasks Completed</i>	<i>Findings/Recommendations</i>
178	141	100	500+

Table 2: Total Results after 44 Trials of Method

During each thirty minute demonstration, we developed between one and eight tasks (at a rate of approximately one per minute); we administered between one and five tasks (usually administering three of every four tasks developed); and volunteers user testers completed between zero and five tasks (usually succeeding in completing two out of three tasks). By the end of each evaluation, moreover, we had made somewhere between five and twenty-five design recommendations to improve the website, each reflecting usability flaws uncovered by our analysis. On average, therefore, in one thirty-minute test, we were able to develop approximately four tasks; we were able to administer approximately three tasks; the users could successfully complete approximately two tasks; and we could make somewhere between ten and fifteen recommendations for design (see Table 3).

<i>Tasks Developed</i>	<i>Tasks Administered</i>	<i>Tasks Completed</i>	<i>Findings/Recommendations</i>
4.0	3.2	2.3	10-15

Table 3: Average Results of one Thirty-Minute Demonstration

These results were, to us at least, simply astonishing. In preparing this method, we had hoped that in thirty minutes we would be able to illustrate the process and perhaps make a handful of observations about usability issues observed in the demonstrations. Given our limited time frame, we could never have imagined the rich information we received (we had even worried we would not uncover any findings worth reporting). When we compare our expectations with our results, our data look almost too good to be true (we have every sympathy with the skeptical

reader)! In a companion paper (in progress), we are exploring from a theoretical perspective why this method is so productive, seemingly against all odds; in the meantime, we urge others to try and replicate the method and see for themselves, while we address below the power of this method as a demonstration tool in practice.

The important question for this paper is whether the *Usability@90mph* method is useful for both demonstrating the user testing process and convincing audiences that user testing is a valuable and efficient way of finding problems with any given interface. To meet our performance/pedagogical goal, we needed to create a significant number of teachable moments in the shortest possible time in each evaluation. We believe there are three factors that indicate that we successfully achieved this goal in each of our 44 tests.

Each evaluation demonstrated many different aspects of the user testing process. The above data analysis shows clearly that we thoroughly demonstrated all parts of the user testing process (from representative task development through recommendations for design improvements) in a very small amount of time. Compared to the amount of time these steps typically take in experimental user testing labs, these numbers were tremendous indeed. Audience members who watched only one evaluation, on average, observed the equivalent of more than fifty hours of user testing and analysis, compressed into thirty minutes.

There was no instance of zero findings in any of the 44 evaluations. At a minimum, we always found at least five to ten usability flaws which we were able to translate into recommendations for design improvements. At no time did we have nothing to say, nor were these stock responses we kept in reserve from some kind of prior evaluation; there were no prior evaluations conducted of any of the sites. All of these findings were specific recommendations that derived directly from the results of the tests.

Participants in the method personally validated the results of the tests in each evaluation. While we cannot claim, given the nature of the method, that these findings were in any way comprehensive or indicative of the majority of usability problems with the evaluated sites, we could and did attempt to validate our findings with the site representatives, volunteer user testers, and audience members during each test. Our goal in doing so was to determine whether our findings would be taken seriously by observers and participants. We found that, again and again, the participants recognized the usability flaws discussed as serious usability problems and agreed that implementing the design recommendations would in fact improve the usability of the site being evaluated. In addition, participants (as well as the evaluators) frequently validated the results of each test by noting how particular findings were often observed in other websites and widely reported upon in the literature on website usability.

Thus, in every single instance, we provided a relatively complete demonstration of the user testing process, consistently produced non-zero recommendations for design, and all participants (evaluators, volunteer user testers, site representatives, and audience members) confirmed the validity of the findings in terms of usability problems that needed to be solved. Whether or not this method produced the most significant usability problems with the sites being tested is irrelevant; at the end of each demonstration, all that mattered was that each participant or observer agreed that the findings were relevant, the demonstrations were valuable, and the

recommendations for design would result in overall usability improvements. The fact that the audience and participants were convinced, allows us to conclude that even when conducted at high speeds, demonstrations of user testing are effective in illustrating the importance of usability analysis.

Success@90mph: Strategies for Successful High-Speed User Testing Demonstrations

This section presents some of the lessons we learned in developing our method that will help others performing similar high-speed demonstrations of user testing. These lessons are divided into four categories by the different roles of participants in *Usability@90mph*.

Evaluators

While evaluators need to have expertise with “regular,” laboratory-based user-testing methods, they also need to understand that *Usability@90mph* is not a “regular” usability method. By definition, *Usability@90mph* occurs very quickly, and the evaluators should practice splitting their focus between macro issues, such as interpreting user actions, and micro issues, such as where the user is clicking. The method will not necessarily run the exact same way each time it is implemented, and evaluators will need to be prepared to make modifications to activities as necessary. In particular, they need to be skilled with scenario-based evaluations (Rosson and Carroll, 2001) and comfortable with quickly developing and adapting scenarios on the fly.

Creating representative scenarios of use at the rate of one each minute can be challenging; evaluators have to pay careful attention to the site representative’s introduction, always thinking of possible tasks for potential users, while avoiding tasks that might be trivially easy or impossible to perform. Developing suitable tasks is easier if the macro-level evaluator has some domain knowledge of the types of interfaces being evaluated. Familiarity with the common usability problems of websites in general or this type of website in particular will also help evaluators develop as many tasks as possible in the shortest amount of time. **The micro-level evaluator should focus on differential analysis: the way in which certain tasks or parts of tasks seem to cause more difficulty or hesitation than others, and the likely causes of those differences due to the underlying site design and the ongoing learning of the site by the volunteers.**

Evaluators should be prepared to make mistakes and recover from them. Having never seen the site before, they cannot expect to develop scenarios that will be consistently suitable for the volunteer test users. Some scenarios may turn out to be impossible to complete, others may be inappropriate given the time restraints, while others still may be problematic because of the evaluator’s incomplete knowledge of the website. The skilled evaluator needs to determine quickly when any given scenario is not working, and move on to other scenarios that may be more likely to yield valuable findings. The more evaluators practice with the *Usability@90mph* method, the better they will get at administering scenarios that address key usability findings and illustrate the value of user testing.

Site Representatives

The *Usability@90mph* method works best when the site representatives are fully involved in the process. Site representatives occasionally had a hard time resisting interfering in the tests; there is a reason, after all, that most usability labs have programmers and designers sitting behind sound-proof windows. During some tests, site representatives could not resist interrupting the test and shouting out comments, whether from frustration or excitement.

The site representatives who learned the most about their site appeared to be those individuals who watched attentively and virtually debugged their system as the user test occurred. The site representatives who saw this method as a learning opportunity appeared to find the results far more useful than site representatives who were very defensive about their site and unwilling to accept constructive criticism. The more frequently site representatives observed the method in action ***beforehand***, the more they seemed to learn when their own site was evaluated.

Finally, the site representatives often had difficulty limiting their site introductions to four minutes. The usability evaluators have to be prepared to move them along, encouraging them to hit the highlights so that the audience (and evaluators) understood the site's purpose without getting bogged down in too many details.

Volunteer User Testers

When running a method such as *Usability@90mph*, the guidelines governing when evaluators should intervene in user tests are very different from laboratory-based user testing. If our volunteer user testers appeared stuck, or we felt that too much time was being spent on a scenario unlikely to return useful results, we would usually intervene and ask them to move on to the next scenario. We often intervened to ask our test users probing questions (a form of cognitive diagnosis) to demonstrate to the audience how relevant questions (e.g., what do you think about what just happened?) can help elucidate more information. Given the extremely limited time available for analysis, we frequently paused the testing process and analyzed some findings on the fly, especially when we were finding many usability problems and worried that we might not be able to cover them all in the last stage of the method. These kinds of interventions are more acceptable in a demonstration method than in laboratory-based user testing, since our primary goal is to create a positive experience for the audience.

Evaluators must also bear in mind that their user testers, as audience volunteers, may not be truly representative users of the interface being evaluated (and certainly not the sort of users one would normally advocate for user testing). In our trials of this method, the audience members were primarily museum professionals, sometimes even museum website designers, and we felt it important to make it clear to the audience any possible limitations this might create. Our volunteers, while unfamiliar with the websites being evaluated, were likely far more web-savvy and far more knowledgeable about museum websites overall than the general population. Despite this, they still experienced many usability problems with the museum websites we evaluated, and a result we often argued that any problems they encountered would likely cause even more trouble for the average museum website visitor. By focusing on the confusions and errors encountered by the volunteer user testers, evaluators can identify problems worth considering in later tests from the perspective of more representative users.

Finally, our volunteer user testers ~~often~~ ****sometimes**** became nervous, and we had to keep reminding them that we were testing the website and not them. Completing a user test in public can be ~~extremely~~ daunting, especially for professionals who might feel particularly embarrassed about making mistakes in front of an audience of their peers. To mitigate this anxiety, in addition to the conventional reassurances given to user testers, evaluators should endeavor throughout the testing process to emphasize the informal, educational, and playful nature of the experience. The scenarios we provided our volunteer user testers typically involved asking them to assume the roles of non-expert users; therefore, any mistakes or other misunderstandings would not reflect on the volunteers' professional status. The public nature of the demonstrations also meant that most volunteers had already seen the testing process at least once and so fully understood the aims and nature of the demonstrations before volunteering. In addition, a feeling of nervousness helped our volunteers understand what it is like to participate in a user test; after having participated in a session, our volunteers frequently said they would be more likely to empathize with frustrated users in the future.

Audience

One difficulty with demonstrating user testing in front of an audience stems from the potential for audience members, either individually or as a group, to interfere with the test in progress. The atmosphere in the room was often like that of a game show, with audience members gasping, laughing, or commenting along with the volunteer user testers. Naturally, such feedback, whether implicit or explicit, can have some effect on the usability test, and as evaluators we sometimes had to remind audience members not to interfere too much with the test in progress. Some interference is a given, however, and usability evaluators must take that into consideration when implementing this method. In addition, even if no one in the audience is making a sound, a room full of people silently watching a user test in action can be extremely intimidating for the volunteer user testers. Evaluators can help volunteers feel safer and less intimidated by creating an atmosphere where the user testers feel like the audience is working with them to evaluate the interface, and not working against them, secretly willing them to fail at the representative tasks.

By asking audience members to assist the evaluators and testers in finding usability problems with the interface, we often uncovered many more problems than we might have if the audience had not been present. Also, audience members often made many good suggestions for representative tasks that identified key usability issues. This level of involvement on the part of the audience helped develop the evangelical nature of *Usability@90mph* as a demonstration method. As audience members spread the word about the effectiveness of user testing, our audiences at these conferences would frequently swell until every seat was taken; often audience members told us that they thought they would stop by for just a few minutes and ended up staying for hours, so riveted were they by the proceedings taking place. Finally, many audience members told us that their involvement in the user tests made them even more likely to stress the importance of user testing in their own organizations than if they had been merely passive observers.

Implications@90mph: Conclusions, Limitations, and Future Directions

Our study of the *Usability@90mph* method indicates that high-speed user testing demonstration methods can be very effective for both demonstrating the user testing process and illustrating the benefits of user testing in general. The fact that we were able to implement this method forty-four times in only twenty-two hours to audiences numbering in the hundreds stands as testimony to the ability of *Usability@90mph* to influence people and increase the visibility of user testing methods. We firmly believe that this method has the potential to demonstrate the capabilities and benefits of user testing quickly and easily, illustrate the importance of user testing to a variety of stakeholders, reach audiences far larger than possible in a user testing laboratory, and convince observers that user testing is something that should be performed early and often. There are three limitations of this demonstration method, however, which are important to point out.

First, the goal of this demonstration method was not to train audience members in user testing methods per se, but rather to educate them about the value of user testing in general. Not everyone has the education or experience necessary to conduct user tests, and we would be dismayed if anyone, after watching one short user testing demonstration, considered him or herself to be an expert at usability analysis! The benefit of this demonstration method was that it allowed audience members to gain some familiarity with the user testing process in its entirety without having to invest one, two, or three days in attending workshops or training sessions. It is our hope, then, that audience members will take their new found appreciation and understanding of the value of user testing back to their organizations to spread the word that user testing is a “good thing” and worth pursuing.

Second, one should bear in mind that conducting successful public demonstrations of user testing is not easy; doing this well takes a great deal of familiarity with experimental user testing methods as well as the ability to apply the theories of usability on the fly to obtain practical results in a very short period of time. We hope that the “lessons learned” provided in this article will help usability evaluators understand how the public and pedagogic nature of our goals in developing *Usability@90mph* sometimes conflicted with the goals of “regular” (i.e. non-demonstration) user testing. The pragmatics of conducting user tests with a focus on education, user advocacy, and performance in front of sizeable audiences naturally constrained the user testing process in some ways, and it is important that the constraints discussed above be firmly understood by anyone attempting to conduct public user testing demonstrations on their own.

Third, the method described in this paper should not be mistaken as a substitute for traditional, experimental, laboratory-based user testing. ****Its purpose is to be a theatrical trailer, not a replacement for the main feature.**** One thirty-minute usability evaluation can never uncover all, or even the majority, of the usability problems with any given interface; it is even possible that the most serious usability flaws will be overlooked by this method. It would, therefore, not be advisable for unskilled evaluators to attempt to substitute the informal method described in this paper for more scientific, formal evaluation methods. In our forty-four trials of this method, we often had to explain to our audience why we felt our results were valid in each circumstance, and in doing so we drew upon many years of personal experience with user testing. Arriving at these sorts of conclusions is not easy, and not just anyone will be able to implement this demonstration method; there could be serious consequences that backfire on the unskilled evaluator.

Nevertheless, with these caveats in mind, we believe that *Usability@90mph*, as a method of raising awareness of and interest in user testing, is replicable by others who have experience with user testing and usability analysis. While the authors have considerable experience in evaluating museum websites in particular, this method does not require specific knowledge of particular types of websites as much as it requires knowledge of interface design and website usability in general. Moreover, the fact that the method requires little advance preparation makes it likely that this approach could be easily adapted for use in other venues or at smaller scales (around a conference table, for example, rather than on a stage). By demonstrating the potential value of user testing to many different types of audiences, we hope to encourage a greater affinity for frequent user testing, and to convince usability educators and evaluators that high-speed demonstration methods like *Usability@90mph* are worth developing and pursuing.

In conclusion, we believe that high-speed, public demonstrations of user testing are an excellent way of encouraging a “culture of usability.” We feel that the results presented above clearly document the power of *Usability@90mph* to demonstrate, quickly and easily, the value of user testing. While our personal tests of this method have been extremely encouraging, a far stronger test would lie in the successful replication of this approach by other usability evaluators or educators. One of our main reasons for publishing this paper, therefore, is to invite others to adopt or adapt our method for their own needs, comparing our findings with their own experiences demonstrating user testing as a combination of evaluative and pedagogic techniques.

References

Kirsten Bauersfeld and Shannon Halgren, 1996. “‘You’ve Got Three Days!’ Case Studies in the Field Techniques for the Time-Challenged,” In: D. Wixon and J. Ramey (editors). *Field Methods Casebook for Software Design*. New York: Wiley, pp. 177-196.

Scott Berkun, 2003. “Interactionary: Sports for Design Training and Team Building,” at <http://www.uiweb.com/dsports>, accessed 9 May 2005.

George Donahue, 2001. “Usability and the Bottom Line,” *IEEE Software*, January/February 2001, pp. 31-37.

Joseph Dumas, 2002. “User-based Evaluations,” In: J. Jackie and A. Sears (editors). *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., pp.1093-1117.

Janice Fraser, 2002. “The Culture of Usability: How to Spend Less and Get More from your Usability-Testing Program,” *New Architect*, August 2002, at <http://www.newarchitectmag.com/documents/s=2450/na0802b/>

JoAnn T. Hackos and Janice C. Redish, 1998. *User and Task Analysis for Interface Design*. New York: John Wiley & Sons, Inc.,

Paul F. Marty and Michael B. Twidale, 2004. "Lost in Gallery Space: A Conceptual Framework for Analyzing the Usability Flaws of Museum Web Sites," *First Monday*, volume 9, number 9 (September), at http://firstmonday.org/issues/issue9_9/marty, accessed 9 May 2005.

Jacob Nielsen, 1994a. "Guerilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier," In: R. G. Bias and D. J. Mayhew (editors). *Cost Justifying Usability*. Boston: Academic Press, pp. 242-272.

Jakob Nielsen, 1994b. *Usability Engineering*. Boston: Academic Press.

Jakob Nielsen, 1999. *Designing Web Usability: The Practice of Simplicity*. Berkeley, CA: New Riders Publishing.

Jakob Nielsen and Robert L. Mack, 1994. *Usability Inspection Methods*. New York: John Wiley & Sons.

Mary Beth Rosson and John M. Carroll, 2001. *Usability Engineering: Scenario-Based Development of Human-Computer Interaction*. San Francisco, CA: Morgan Kaufmann Publishing.

Daniel Rowley, 1994. "Usability Testing in the Field: Bringing the Laboratory to the User," In: *Proc. of CHI-94: Human Factors in Computing Systems*, pp.252-257.

Jeff Rubin, 1994. *Handbook of User Testing*. New York: John Wiley & Sons.

Ben Shneiderman, 2000. "Universal Usability," *Communications of the ACM*, volume 43, number 5, pp. 84-91.

Ben Shneiderman, 2002. *Leonardo's Laptop: Human Needs and the New Computing Technologies*. Boston: MIT Press.

David A. Siegel, 2003. "The Business Case for User-Centered Design: Increasing your Power of Persuasion," *Interactions*, volume 10, number 3, pp. 30-36.

Bruce Thomas, 1996. "Quick and Dirty Usability Tests," In: P. Jordan, B. Thomas, B. Weerdmeester, and I. McClelland (editors). *Usability Evaluation in Industry*. London: Taylor & Francis, Ltd., pp.107-114.

L. Van Waes, 2000. "Thinking Aloud as a Method for Testing the Usability of Websites," *IEEE Transaction on Professional Communication*, volume 43, pp. 279-291.

Daniel Wildman, 1995. "Getting the Most from Paired User Testing," *Interactions*, volume 2, number 3, pp. 21-27.

Dennis Wixon and Judith Ramey (editors), 1996. *Field Methods Casebook for Software Design*. New York: John Wiley & Sons.

About the Authors

Paul F. Marty is Assistant Professor in the College of Information at Florida State University.
Email: marty@fsu.edu

Michael B. Twidale is Associate Professor in the Graduate School of Library and Information Science (GSLIS) at the University of Illinois at Urbana-Champaign. Email: twidale@uiuc.edu