

# Chapter 8

## Usability Tests

Jerilyn R. Veldof

A good interface should get out of the way of the learner. In fact, it should be nearly invisible, allowing the learner's complete focus to be directed on the content. If we fail this challenge—if the Web navigation is confusing, labels obscure, framework incomprehensible, and screens dense—we may lose our audience before they begin to learn. Interface usability is, therefore, essential to the development of effective Web-based instruction products.

“Know thy user, for they are not you” is an oft-cited maxim repeated by many usability experts. (Dodge, 2000, Mandel, 2000) You probably know a great deal about your audience either from personal interaction or your research in the planning stages of the project. Despite this, there can still be a wide chasm separating the way librarians think about both research and Web-based instruction and the way students think about those things. In a study conducted during the development of the online tutorial at the University of Minnesota (U of M) we labeled this distinction the “clash of the mental models” (see Figure 8-1).

When we at the U of M designed our Web-based tutorial to fit librarians' mental models, we were unwittingly setting our target audience (lower-division undergraduates) up for failure. One of the keys to designing a usable Web site, then, is gathering information about the target audience and the way that they interact with the site or product.

**Figure 8–1. Clash of the Mental Models**

LIBRARIANS	STUDENTS
Research is an end.	Research is a means to an end.
Good research is valued.	Good grade is valued.
Research is complicated.	Research is fast and easy.
Research is something to learn how to do.	Research is simply something you do.
Learning is preferred through books and pathfinders.	Learning is preferred through games and interactive simulations.

(Veldof and Beavers, 2001)

### USER-CENTERED DESIGN

The degree to which a Web site is usable is in direct proportion to the degree that it is user centered. User-centered design challenges us to create a design that fits the user—not to make the user fit the design. With this approach, we shift from designer-centered questions such as “Does it work for me? Do I like it? Do I find it easy to use?” to user-centered questions such as “Does it work for the learners I’m designing this for? Do they like it? Do they find it easy to use?”

This is a significant shift. Many library Web sites have been developed “behind closed doors” with input only from librarians and technical staff. The results are often text-heavy, jargon-filled sites that require users have prior experience with libraries and online research. Ultimately, these Web sites create barriers between the audience and the content, increasing users’ frustration and failure rates.

User-centered design, on the other hand, provides the project team with evidence that their design works before it ever gets released to the public. The best way to ensure a functional design is to involve members of the target audience in the process from the beginning. Users should help define what should be on the site and offer the project team feedback on the effectiveness of the site. Usability methods assist project teams in

obtaining this valuable input. The Usability Methods Toolbox (<http://jthom.best.vwh.net/usability/>) includes the range of usability methods including heuristic evaluations, cognitive walk-throughs, mix-and-match tests, and card sorting. Dickstein and Mills illustrate use of these techniques in a library setting (2000). All of these methods can be useful in a user-centered design process but, ultimately, the usability test is the most telling.

In the simplest terms, a test for usability involves observing a user as he or she navigates through a Web site and taking note of where they run into trouble. These observations allow you to fix problems before the site is released to the public.

Many products that we buy have been usability tested with customers. Large companies, such as Target Corporation and Microsoft, have in-house usability labs while smaller sized companies may hire outside usability experts to conduct testing for them in their own facilities. Computer usability labs have rooms with one-way mirrors and cameras inconspicuously placed to capture keystrokes, screens, and facial reactions. The room behind the one-way mirror might have a video capture board and several monitors so the project team and usability experts can observe. When users are brought in to test a product, no reaction, no move, no comment, goes unrecorded. The result is well-documented analysis and feedback for the project team.

Most likely, however, your library or institution is not blessed with access to such a facility or with the experts to conduct your testing. Because most project teams have limited budgets and restricted timelines, the usability process explained here is stripped down to the most basic, effective elements.

## ITERATIVE DESIGN PROCESS

Your Web-based instruction site should be tested with users throughout the entire process—from planning through evaluation. Project teams will often test with users, build parts of the site, test with more users, and rebuild, many times during development.

In the planning stage, the project team might choose to test products similar to the one they are designing. Usability testing of these sites provides the team with an opportunity to learn

from the best and avoid the worst. Questions that the project team has about the site architecture, metaphors, interactivity, labels, language, and tone can be addressed during this initial round of testing. At U of M, for example, we initially tested up to nine other Web sites in fifteen-minute rotations. Three tests at a time were held simultaneously in a reserved computer lab. In this way within hours we were able to quickly test how well elements of other Web sites performed before borrowing or adapting these elements in the first draft of our design.

In user-centered design, the initial design is often created and tested through basic paper prototypes, a technique where the project team sketches out designs on paper to test with users. The team may create several versions of paper prototypes that are tested and improved until they reach a level of usability that justifies the time and expense to move the design online. The computer prototype may not be fully functioning initially—perhaps just a graphical design without the programming behind it. During early usability tests, the test moderator may move between the computer interface and a paper prototype for different layers of the site. These tests are usually brief since the team will have few layers of the site developed at this early stage.

Once the team is satisfied that they have a usable prototype and that the basic functionality of the site appears to be effective, they may invest in some programming and interface design. These iterations are again tested with members of the target audience until the team has enough evidence that learners are able to successfully use the site.

## ROLES AND RESPONSIBILITIES

There are generally three to four people present at each usability test: a test monitor, a recorder, an observer, and a test participant.

### **Test Monitor**

The test monitor is the person who administers the usability test and may also be the person who writes down the path, or se-

ries of clicks, the user takes throughout the test. Usability experts warn that the people creating the product should not do the testing. It is extremely hard to remain objective if one is extricably connected to the product. Test monitors may say rather innocuous things like, "Wasn't that difficult?" and "That looked easy. Was it?" that are in fact leading questions and reflect their own personal biases about the interface. Bias will discredit the test results, potentially turning the entire testing process into wasted time. Recruiting someone not involved in the design to be your test monitor may be the easiest low-cost option to reduce test bias. If that person is also developing a site, you might be able to trade test monitor responsibilities for each other's projects.

If you work in an institution where others are unfamiliar with usability testing, the project team may need to lead the effort themselves. Choose a test monitor on your team who can be most objective and least defensive about the product. Including time at the end of each test for the recorder and the observer to give the test monitor feedback may help reduce obvious testing bias. Jeffrey Rubin's *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests* has an excellent chapter on conducting usability tests. (1994: 213–242)

### Recorder

The recorder documents everything the user is saying—verbatim if possible. The verbatim record of the test helps the project team to separate out what they think they heard with what was actually said. Do not rely on memory alone. It is easy to remember only the things that you want to hear, that you expect to hear, or that support your beliefs about the design. After the test, return to the notes to make sure your biases did not overpower the test participant.

Another benefit of the verbatim record is that it provides the project team with the specific language of the test participants. This language may be exactly what you need to improve usability in the next iteration's labels, descriptions, text, and help pages.

### **Observer**

The observer pays close attention to the reactions of the test participant. While the test monitor and the recorder are often busy with their responsibilities, the observer has no other task than listening and watching the participant. Before the test is over, they may notice certain nuances and perhaps even develop possible solutions. At some point, each of the project team members should play this testing role.

One of the greatest benefits for allowing observers into testing is that they often become strong advocates for the usability of the product. Even the most obstinate person is likely to have their assumptions jarred when they witness students struggling and failing with the site design. There is no better way to build support for iterative user-centered design.

### **Test Participant**

Test participants should be people who reflect the demographics and characteristics of the target audience. If your target audience includes freshmen in introductory composition courses, then ensure your test participants are freshmen who haven't yet used the library. If your target audience includes retired senior citizens learning to use the Internet for the first time, find individuals who fit that profile.

Typically, it can take anywhere from five to ten test participants per design iteration to uncover the majority of a site's problems. Jakob Nielsen points out that "As you add more and more users, you learn less and less because you will keep seeing the same things again and again. There is no real need to keep observing the same thing multiple times . . ." (Nielsen, 2000)

Do not "reuse" test participants from iteration to iteration. They will become too familiar with the intent of the product, the navigation, the labeling, and will build expectations for what they are able to do at the site. Recruitment of participants will be discussed below.

## PREPARING FOR AND CONDUCTING USABILITY TESTS

Now that you know who will be involved in testing, let's talk about how to prepare for and conduct the test.

### **Step 1: Identify tasks to test.**

A usability test is made up of tasks that the participant is asked to complete. What is tested is often what gets improved, so it is important to choose tasks wisely. Focus groups and a review of the goals and learning objectives for the site are essential for task identification.

Focus groups should be designed to identify 1) reasons the learner might use the product and 2) goals they might want to accomplish with it. Focus groups might be conducted with several stakeholder groups. For example, an online tutorial might be designed to augment faculty's own instruction on library research, so a focus group of these faculty and their expectations for the product would be very helpful. A focus group with the end user of the tutorial would round out the information gathered from the faculty focus group. Responses and discussion in the focus groups should be recorded verbatim. Often the exact wording for tasks can come from these transcripts.

The project team should also examine the site's goals and learning objectives developed in the planning stages. Why are you designing the site? What do you want your learners to learn? Do the learners' needs and expectations dovetail with the designers' expectations? Based on these responses and focus group data, begin to make a list of the key tasks.

Note that with Web-based learning, tasks change in relation to the way the learner will be using the learning product. For example, the designers of the U of M's information literacy tutorial learned through focus groups that most undergraduates would not take a tutorial voluntarily. The primary audience for the tutorial became, therefore, students who are required by their instructor to complete the tutorial. A relevant task for this audience might

be: "You have been assigned to complete Module 3 of the library tutorial, take the quiz at the completion of the module, and send the results to your instructor. How would you go about doing that using this Web site?" This is a "real" user task as opposed to a task that might say, "You're trying to find articles in the library. How might you do that using this site?" since that is not a task that the target learner would be completing in the actual product.

Task questions should be piloted with members of your target audience for clarity. It is easy to write a question that tests something unintended, or to confuse a test participant with an unclear question. Piloting questions also allows the project team to determine approximately how long an actual test will take and to adjust incentives accordingly.

### **Step 2: Build a post-test questionnaire.**

A questionnaire given after the test is completed is sometimes a good way for introverted participants to provide thoughtful feedback in a relaxed manner. It gives the participant time alone without the monitor, recorder, and observers hanging onto every word. The downside to including a post-test questionnaire is that someone has to compile the data and analyze it. Depending on how quickly the iterative design process is going, this could needlessly slow down the group. One way to address this is to add open-ended debriefing type questions to the end of the verbal part of the test such as:

- What did you think about the site?
- What were the best features of this site?
- What would make it easier to use this site?

These kinds of questions may elicit some of the most insightful comments from participants and lead to dramatic design improvements.

### **Step 3: Choose and train the test monitors.**

There are a number of ways that a monitor can ensure that the test goes smoothly, just as there are a number of ways a monitor can sabotage a test. Monitors should be trained in the practice of administering the test and have a chance



to conduct pilot tests. Other team members can provide feedback and guidance to the trainee. Rubin (1994) wrote an excellent guide useful for training test monitors, in which he lists and elaborates on the key characteristics of a good test monitor. Although it is true a monitor can be trained, identifying individuals who may be predisposed to being a successful monitor may help speed the process along.

Characteristics of a good test monitor include:

- Grounding in the basics of usability engineering.
- Quick learner.
- Instant rapport with participants.
- Excellent memory.
- Good listener.
- Comfortable with ambiguity.
- Flexibility.
- Long attention span.
- Empathic “people person.”
- “Big picture” thinker.
- Good communicator.
- Good organizer and coordinator. (Rubin, 1994: 67–71)

#### **Step 4: Recruit test participants.**

Unless you have administrative staff who can schedule test participants, you might leave recruitment until you are ready to administer the test. Try recruiting participants from areas where they congregate, such as dining halls and study areas in the case of students, or shopping areas and food courts in the case of the public. Setting up a laptop or using the corner of a nearby computer lab may be sufficient. Only recruit users from within the library if current library users are your target learners. Otherwise, push yourself to go where your target audience is—outside the library.

For tests over 10 or 15 minutes, you will most likely need to offer incentives for participation. Students may be motivated to participate with as little as \$5 for a 15-minute test, or \$10 or \$15 for a full hour. A free-lunch voucher, gift certificate, free copy cards, or private one-to-one instruction

session after the test might motivate some people (such as faculty). Experiment with incentives and recruiting locations until you find a good formula that matches your audience's interests and needs.

Some institutions will require that the group responsible for testing complete a report outlining the usability test and testing procedures in order to comply with human-subject testing guidelines. You may also be required to have your test participants sign a special waiver developed by an attorney. These requirements will vary by institution.

### **Step 5: Administer the test.**

The test monitor will be the main contact with the participant during the test. He or she should be sitting in such a way that they can easily make eye contact with the participant but also see the computer screen. The recorder should be sitting out of the sight of the participant but still in easy view of the screen. Observers should sit as far away as hearing and viewing permits. All observers and the recorder need to refrain from any comments, sighs, laughs, or gasps during the test (and honestly, this is very difficult to do!).

The monitor should introduce himself or herself, the recorder, and the observers to the participant and work to put that person at ease. Emphasize that this is not a test of the participant's skills or knowledge, rather a way to find out what is working or not working with a Web site that the library is designing. Notice that you do not want to put the participant in the position of criticizing you as a designer, so try to distance yourself a bit from the product.

Participants are then asked to think aloud while completing the test. The think-aloud protocol sounds almost like a stream of consciousness from the participant: "Oh wow, there's a lot to read on this page. I have no idea where to go. Oh, maybe I'll try clicking on this link here or, no, no, I'm going to try this one—I think this is the right one . . ." The monitor might gently interject with, "Why do you think that's the right one?" And the thinking aloud would continue. Whenever the participant lapses into si-

lence the test monitor's job is to draw the person back out. The monitor might ask questions like, "What are you thinking right now? Where are you looking? What are you looking for? What did you think should have been here on this page? Where were you expecting that link to go?" Questions need to be extremely neutral and non-leading.

#### **Step 6: Administer the post-test questionnaire.**

Some participants are harder than others to coax into the think-aloud process. Introverts may appreciate some time to mull over questions and provide some feedback without the stress of the actual test. Let participants take the post-test questionnaire in a quiet area where the group is not debriefing.

#### **Step 7: Debrief with the group.**

At the end of each test, schedule time for a short debriefing while your thoughts are still fresh. Capture the key observations that you have consensus on and the key changes that you think address the problems. You might review the transcripts from the recorder to address any conflicting recollections. At this point there may be a few obvious problems that you already agree need to be fixed. After all testing in a particular round has been completed, meet again to compare across tests. Are there recurring themes? Are most test participants getting confused in the same places? Does the entire site organization need to be rethought or would it be enough to improve the language and relabel the links? What kinds of pointers can you place around the site in case learners get lost?

#### **Step 8: Make changes.**

Depending on how far into the iterative design process you are at this point, you might now be able to sketch out a new design, or if further in the process, you might create a list of specific changes for the Web interface designer or programmer.

#### **Step 9: Start again.**

The design team should decide before even starting how they will define success for the site. Initial testing with

other products can help the team define reasonable benchmarks. For example, if testing shows that 60 percent of participants are successful completing a particular task in a similar product, the team might decide to set their target at 80 percent. During each iteration, the team would capture whether or not the participant was able to complete the task on his or her own before giving up. As soon as the target of 80 percent is reached, the team could then focus on other problematic aspects of the site. Alternatively, success is sometimes defined by the time it takes to complete a task. You might set a goal that a task should only take a certain amount of time to complete. If a participant takes that amount of time or less, the site would be deemed successful for that particular task.

Regardless of how you define success, taking the time to define your success measures may pay off in the end. I have yet to see a Web site without usability problems. Even if you undergo a dozen iterations of design and testing, your site will still cause confusion for some learners. But at some point you are going to have to stop testing and redesigning. Having set success measures will allow you to report to the library that you have good indication that "x" percentage of your users will be successful, that this was your goal, and that the product is therefore ready to be released. Case closed . . . until the next redesign.

What drives a whole new redesign cycle? There might be a number of factors that point to a need for a redesign:

- Availability of new technology that would improve the effectiveness of the site.
- Wider use of newer technologies by users, allowing you to upgrade your site.
- Shift in the goals of the project, use of the site, or nature of the target audience.
- Higher expectations from your target audience over time.
- Outdated or old-fashioned design compared to other sites your target audience visits.

## CASE STUDY

In the fall of 1998, the U of M conducted usability tests focusing on the online tutorial "QuickStudy: A Library Research Guide." Twelve undergraduates completed a test consisting of eight tasks and four follow-up questions. After the first round of testing it became apparent, particularly from student comments, that the team had designed task questions unrealistic for the target audience. The project team reworked all test questions, redesigning them to reflect how students would use the site. The new task questions were reduced in number, although the time spent on each task was more substantial.

Some questions focused on specific tasks such as:

You're taking a Biology course on environmental management and your final project is to write a ten-page paper. Your professor has asked each of you to come up with a great topic. Practically the whole class is made up of freshmen and you're all unsure about how to create a good paper topic. Your professor tells you to use something called QuickStudy and includes these directions in your syllabus [student is shown the following]:

- a) Go to QuickStudy: Library Research Guide at: <http://tutorial.lib.umn.edu>.
- b) Choose and complete "Starting your Research"

Now, let's say you're sitting in a computer lab ready to start this assignment. What do you do?

Other questions asked participants to consider the site more generally after they had completed the task questions:

- What are your general thoughts about this tool?
- What did you think about how the site is arranged?
- What are your thoughts on navigating the site?
- Specifics (anything the test monitor wants to follow up on from the test).
- What were the best things about this site?
- What things you would like changed on this site?

- Would you ever use this again? Why or why not?
- Questions from observers.

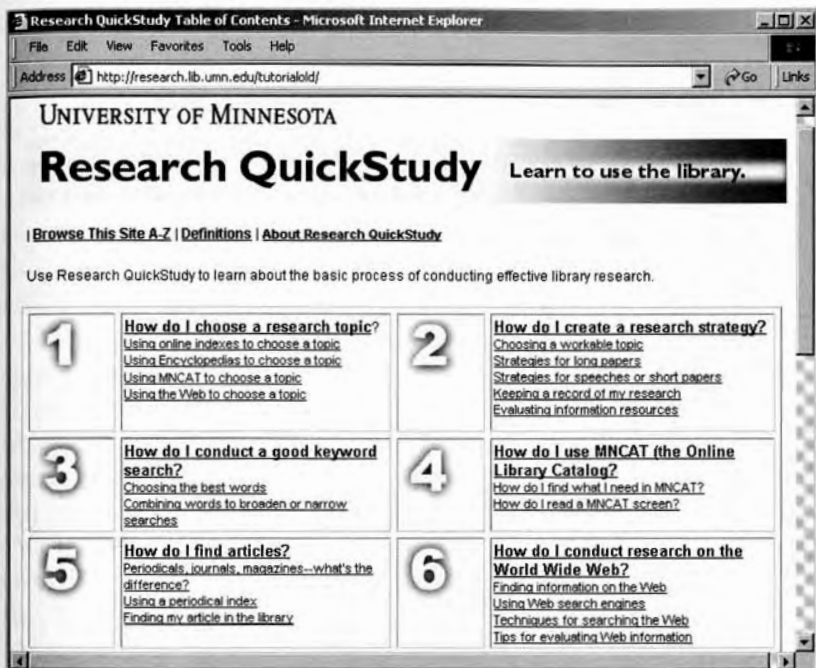
Undergraduate students completing a project in a usability class conducted a third round of testing of the QuickStudy tutorial. The class first met with the project team in the library to discuss the goals for testing and then devised a test script, arranged and administered testing, and provided a report to the project team. The ability to observe the tests and ask follow-up questions of the participants helped make this experience useful to the project team. Veldof and Beavers offer a more complete discussion of a range of the testing methods used throughout the process. (2001)

User insight and feedback shook up many assumptions about the learners and their use of the site. Comments helped inform nitty-gritty decisions about navigation, labels, tone, content density, page length, and instructional approaches. Examples below reflect some changes made to early design iterations:

- Removed elements of the book metaphor, such as “table of contents.”
- Removed a search button that implied the product was a search engine.
- Added formatting such as spacing, bolding, and bullets in recognition that users skim content.
- Repositioned links at the bottom of pages with warnings that they would lead out of the tutorial.
- Added guided exercises for more hands-on learning.
- Added as much interaction and application as programming and time permitted.

Over time the team got closer to a product that students proved was relatively usable. Figure 8–2 shows the homepage of the older version of the tutorial and Figure 8–3 shows the homepage of the redesigned site. Although the project team was aware that other changes could be made to design a more effective learning environment and to improve usability, the team was confident that the redesigned product was reasonably effective and would ensure the site had a shelf life for another two to three years.

**Figure 8–2. Initial Version of QuickStudy Tutorial (1997).**  
 URL: <http://research.lib.umn.edu/tutorialold/>

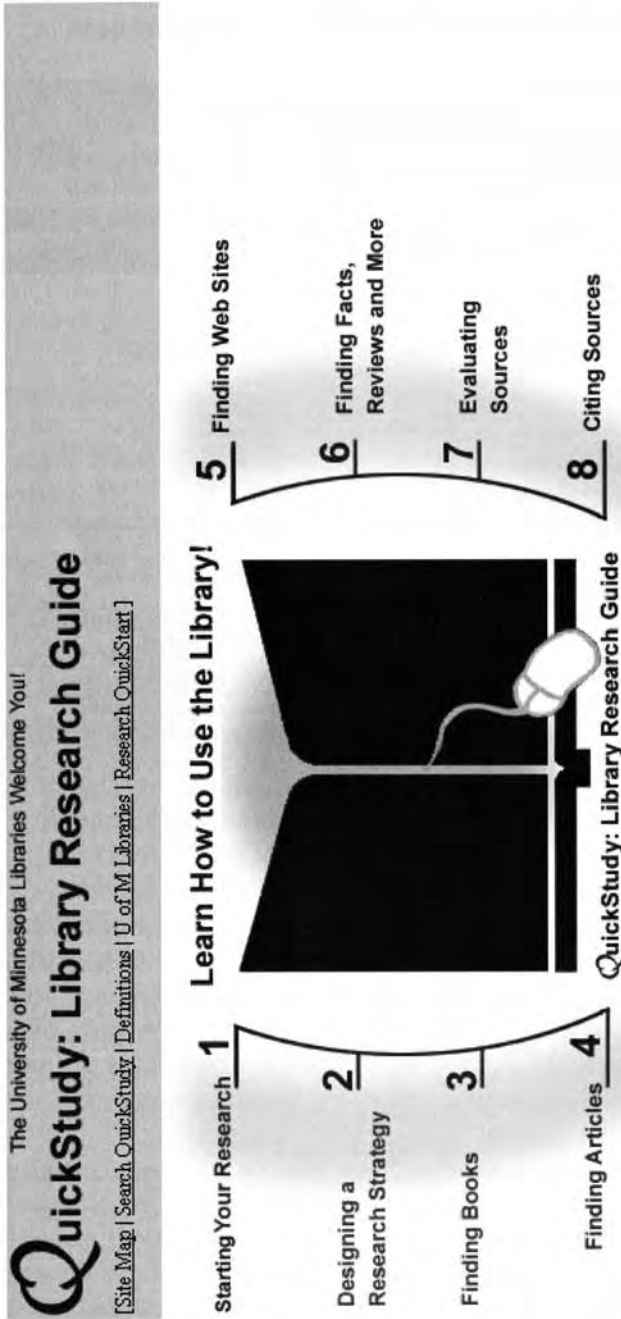


## CONCLUSION

Usability testing will uncover just how challenging it is for our users to effectively navigate through our Web sites, including those developed for online education. During testing, you will see users struggle with the most basic tasks, look right over the links they need to choose, and click on other links at inappropriate times. They will express frustration—some will get angry, others teary-eyed.

You, as a designer, now have the tools to reduce that frustration and anger by making your Web-based instruction site as usable as possible. Sometimes just changing the names of some links, moving buttons to a different part of the screen, and removing clutter from your pages, can make a substantial dif-

Figure 8-3. Revised Version of QuickStudy Tutorial (1999). URL: <http://tutorial.lib.umn.edu/>



[ [Ask Us!](#) | [Contact Us](#) | [For Instructors](#) | [About QuickStudy](#) ]



ference. There is so much that can be done to help make the user successful. Even if you can't figure out how to remove a barrier, or how to redirect users to the "right" place on the page, create safety nets where your users tend to stray so that, as far as they know, everything's going smoothly.

If you're still looking for the silver bullet and asking "What is one thing that I should do if I want to make sure my site is easy to use," heed usability consultant Steve Krug's advice. "The answer is simple," he says. It's not, "Nothing important should ever be more than two clicks away," or "Speak the user's language," or even "Be consistent." It's . . . "Don't make me think!" (2000: 11)

Your Web-based instruction product should be so easy to use that your test participants think it's been a waste of their time to be part of testing it. So easy to use that they say to you, "Well, obviously I'd click here." So easy to use that they look at you funny when you ask them at the end of the test what they'd suggest you do differently. You want them to say things like, "Well, maybe you could make this part of the page a lighter color," and not, "It's really confusing. I just don't think I'd use this at all." Users should not have to think about using the interface—they should be thinking about the content. Conducting even just two or three rounds of usability testing could make the difference between a confusing site and an easy-to-use one.

## REFERENCES

- Dickstein, Ruth, and Victoria A. Mills. 2000. "Usability Testing at the University of Arizona Library: How to Let the Users in on the Design." *Information Technology and Libraries* 19, no.2: 144–151.
- Dodge, John. 2000. "Testing, Testing: Firms Look to Science to Design Web Sites." *WSJ.com: The Wall Street Journal* (September 19) Available: <http://cob.jmu.edu/williac/mktg470/dailyupdate/Websites.htm>.
- Krug, Steve. 2000. *Don't Make Me Think! A Common Sense Approach to Web Usability*. Indianapolis, IN: New Riders Publishing.
- Mandel, Theo. 2000. "Usability Myths and Maxims." Available: [www.theomandel.com/tm-resources/myths.html](http://www.theomandel.com/tm-resources/myths.html).
- Nielsen, Jakob. 2000. "Why You Only Need to Test with 5 Users." *useit.com: Jakob Nielsen's Alertbox*. Available: [www.useit.com/alertbox/20000319.html](http://www.useit.com/alertbox/20000319.html).

- Rubin, Jeffrey. 1994. *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. New York: Wiley.
- Veldof, Jerilyn R., and Karen Beavers. 2001. "Going Mental: Tackling Mental Models for the Online Library Tutorial." *Research Strategies* 18, no.1: 3-20.