The Patron Strikes Back:

A review of recent library web site usability studies

P. Jason Morrison

Kent State University

Information Use and Services

14 July 2005

Introduction

Usability is a hot topic in both the academic and business world, with many books, magazine articles, and scholarly journal articles published each year. Web site designers, businesses and other organizations are trying to make their sites as easy to use as possible. Virtually every type of information provider is looking for ways to do so better. Libraries are in an interesting position in relation to the World Wide Web—traditionally the first place for the public (or their particular user community) to go for information, they now find themselves confronted by patrons that want to simply Google for everything. Library web sites can fill important roles in gathering, organizing, and presenting high-quality information, but are library site designers and library science researchers interested in usability? This literature review looks at ten recent (2001 or later) studies of web site usability in library settings to test the water—what kind of usability research is being done, what are the methodologies and measures used, and what are the main results? This review will examine each article in turn and then conclude with comments on the major themes of the research.

Although a general discussion of the definition and history of usability and usability testing is beyond the scope of this review, it has been defined as when "the people who use the product can do so quickly an easily to accomplish their own tasks." (Dumas and Redish, 1999, p. 4) Usability can be assessed in a number of ways. Users can be surveyed, experts can be assigned to review sites, live usability tests can be run with users' attempts recorded and analyzed, general guidelines can be used to audit sites (Hom, 1998), and classical experimentation is even possible. Although usability is a relatively recent field, it is easy to find a number of articles on usability testing in academic journals. Since library catalog and search systems are often created to make it easier for users to access information, it is no surprise that a number of usability studies have been done on them.

This review is concerned with the methodology used in each article. How was usability assessed on a specific web site or application? What were the measures, what data was collected, and how was the test, survey, or study administered? In addition, attention is paid to the literature reviews for the source or inspiration of the methodology. Findings for each article are reported and trends analyzed. Finally, overall themes, dominant methodologies, and interesting exceptions are discussed. This review is not exhaustive, but should provide an illustration of current research.

Reviews

In "Usability testing of an academic library Web site: a case study" Battleson, Booth, and Weintrop (2001) argue that library websites, as information gateways, must fight "information overload" and concentrate on usability. This report details major concepts of usability and usability testing as well as their use specifically in library settings and then describes a case study at the University of Buffalo libraries. The three major goals for user interfaces in HCI, in their view, are providing task support, being usable, and having an aesthetically pleasant design.

The literature review mentions several books, primarily Nielsen's *Usability Engineering* (1993) but also Rubin's *The Handbook of Usability Testing* (1994) and Dumas and Redish's *A Practical Guide to Usability Testing* (1993). Works consulted that dealt more specifically with web site usability included Nielsen's *Designing Web Usability* (1999), Spool's *Web Site Usability* (1997), and Head's *Design Wise* (1999). The authors note that although formal usability testing began with quantitative studies in psychology, they now are more often concerned with rapid, useful results. They also mention a few library specific usability articles by Eliasen, McKinstry, and Fraser (1997), Chisman, Diller, and Walbridge (1999), and Veldof, Prasse, and Mills (1999) although they did not take their methodology from them. Research by Nielsen and Landauer, (1993) is cited as inspiration for the relatively small number of participants used.

Development of the testing methodology began with asking, "who are the users; what must they accomplish; and what support should the site provide?"

User goals were determined by the test committee and web site manager asking themselves what the population would expect to accomplish, and fell under three categories:

- Identifying items in the collections;
- Locating the most appropriate resource to find journal articles on a specific topic; and
- Finding a starting point for researching a topic without little knowledge of specific sources.

Although no actual user tasks were surveyed, testers tried to create tasks that reflected real questions users might be faced with. Based on other studies, tasks were limited to a small number (11), and testers attempted to choose tasks that could definitely be accomplished with the site and did not have too many possible "correct" answers. Rather than employing video recording or eye tracking, researchers asked users to think aloud while navigating. Each test involved a moderator that interacted with the participant and a scribe that recorded navigation and comments. Unlike some of the other tests described in this review, tasks were not limited uniformly by time but instead a number of false tries were allowed for different questions before clarification was given. Other measures recorded included reaction time, body language, and a participant post-test questionnaire and a tester post-test questionnaire. It is important to note the questionnaires asked the opinion of the test process rather than the site being tested.

For the test, undergraduates with little experience were chosen as the test population and 11 were recruited from English 101 classes and offered a gift certificate to a local store for participation. Actual tests were performed in an empty classroom by appointment. Two dry-run tests were run in order to test the methodology.

Some of the measures used were difficult to determine—for example, it was hard to tell if a user had learned something about the site or had simply made a guess. Other measures—specifically those that measured error—were easier to interpret, since there were objectively ideal, satisfactory, and incorrect

answers.

Results indicated that participants had no trouble using the "Libraries Catalog" link to find books, but usually took two tries to figure out that journals were under "Online Resources." Some other labels on lower-level pages such as "Quick Start" and "Reference Resources" were almost totally ignored.

The researchers concluded that the test found problems that would have otherwise been ignored and were impressed by the results and cost effectiveness of the methodology. They felt that the results had shown the validity of the approach.

McMullen reports on testing efforts at Roger Williams University (RWU) to redesign the library web site to better meet user needs in "Usability testing in a library Web site redesign project" (2001). The web site was seen as an increasingly important point of contact with patrons, especially with the rise of distance learning and remote access. Previous designs of the site revolved around the need to add and organize an ever-growing list of resources, rather than the specific needs of the user population, so a user-centered process was needed for the redesign.

The author observed that user success is often determined by user knowledge both from anecdotal accounts of the old web site and Preece's 1994 book *Human-Computer Interaction.* Marchionini's *Information Seeking in Electronic Environments* (1995) is also mentioned, and Nielsen, Spool, Instone, and Lynch are named dropped as influencing the design process (but not the test design specifically). An online "Alertbox" column by Neilson (2000) is quoted to justify testing with a small number of participants.

Unlike similar studies in this review, the report on this redesign went all the way back to project requirements. The site should:

- provide access to the online library catalog;
- provide access to online periodical databases and indexes;
- provide access to online reference materials;
- provide access to web resources and search engines;
- offer information about the library itself;
- provide support; and
- help users identify what they need.

User groups identified included experienced users, such as faculty, and inexperienced users, such as freshman and clerical staff. In order to figure out appropriate user tasks, the researchers began with their familiarity with the types of questions received at the reference desk and created a list of cognitive tasks required to achieve a goal, for example finding a journal article.

This project involved three distinct phases: first, an observation-interview style study was done using the current web site with 14 users of varying experience levels. Participants we asked to think aloud as they completed tasks based on their own, differing information needs while an observer recorded their thoughts and navigation path. Analysis of the results of phase found that users were overwhelmed by the number of choices on the initial interface, did not understand the terminology used, were not interested in reading help or guides before use, were given the same interface for novice and advanced users, and found the site boring and uninteresting. Users reported that they were happy with the interface despite difficulties and even blamed themselves, rather than the interface, when frustrated.

Efforts to redesign the site before the second phase of the test are described in detail, including prototyping, cognitive walkthrough, and heuristic evaluation. The second phase of the study differed from the first in that it pitted two prototype designs against each other, predetermined tasks were assigned to users, and a $5.00 copy card was offered as an incentive for participation. Five students from various majors participated.

In the third phase, a more formal usability test was conducted, again with five users, including a guide for conducting the observation-interview, a pre and post survey, and nine information seeking tasks. Major findings were that users did not want to read web pages, and that they had a hard time locating journal articles—especially when resources were labeled as being "databases." The author also found that users took different paths to find the same resource, which their site supported.

Although many of the studies reviewed involve academic libraries, "Usability study of the MnLINK Gateway" by Roca and Nord (2001) describes a system designed for the public, namely the Minnesota Library Information Network (MnLINK) Gateway. The paper describes usability testing done on the system, an implementation of OCLC SiteSearch. Although the system was designed for the public, this paper details early user testing with college students.

The literature review in this study goes all the way back to the behaviors of information seekers, through earlier studies of online catalog use Park (1997), Peters (1998), and Tolle (1983) through the field of human-computer interaction (HCI). The process used to design and implement the site followed a user-centric process from Abels et al. (1997). The rationale for the paper-prototype testing is from Hackos and Redish (1997), Rudd et al. (1996), Rubin (1994) and Dumas and Redish (1993).

The was designed to asses three things: users' ability to complete research tasks, understanding of navigation, and understanding of the interface. The research was broken into two phases, with the results of the first phase reported here. Phase I was a paper-prototyping test using a large number of users (101) from a convenience sample (students in writing courses at MSU). Participants were given notebooks with tasks with screenshots from the site and asked to identify page elements or choose a link to follow to complete the task. The second phase would be a formal usability study.

Results were clear. Only half of the participants successfully navigated the opening screen, the vast majority could identify the help button but almost none used it, there were more wrong answers than those left blank, ways in which the site differed from web design conventions had lower performance, and users had trouble with library terminology. Self-assessed experience and class level both correlated positively with performance.

Researchers found this method to be a quick and easy way to identify major usability issues but acknowledged the limitations of paper prototypes. For example, users did not have any real feedback from their choices and could not be asked to do more complicated tasks. This and the limited sample lead them to believe further research would be worthwhile.

In "A case study of the usability testing of the University of South Florida's virtual library interface design," Allen (2002) explains that usability testing is becoming a common, useful practice for library web sites. The paper chronicles the usability testing done as part of the redesign of the USF library web site. The site began in 1996 as a central location for library resources with a small number of databases,

but grew considerably (from 20 to over 300). Links to additional resources were largely unused except by librarians. Growth, and anecdotal evidence of user difficulty, led to the redesign.

The author praises Dickstein and Mills' 2000 study of the University of Arizona's library web site. In addition, Rubin's *Handbook of Usability Testing* (1994) and Dumas and Redish's *A Practical Guide to Usability Testing* (1993) guided the methods of this study. A combination of Rubin's assessment and comparison tests was used for this study.

One reason this paper stands out is the way in which the testing method was chosen. Testers were torn between doing an unmediated test with a large number of users or a mediated test with a smaller number of users and more detailed data, so and experiment was run involving small scale tests of both methods. In this case the main difference between the two methods was who recorded the navigation paths of the users – in the unmediated test, participants were asked to self-record. The experimental testing instrument consisted of first 15 tasks and six subjective questions about the design, with users asked to complete the tasks as far as they could within seven clicks. One result of the experiment was a few wording changes in the tasks to make them more similar to user vocabulary, and one question was dropped as it seemed irrelevant. More important to the decision, though, the experiment found the self-recording users left gaps in their paths, sometimes wrote illegibly, gave up on tasks without trying, and took nearly twice as long to complete the tasks. On the other hand, mediators reported that participants were visibly effected by their, looking for clues or approval, and mediators had difficulty recording navigation paths at the rate users moved. In the end, it was decided to record navigation automatically using web server logs, eliminating extraneous information from the logs and recording start and stop time as well as IP address to match participants with logs.

In the actual usability test, two sets of 16 participants were recruited from classes and were mostly new freshman students with little experience with the library web site, although there were some upperclassmen as well. The only incentive was release from one hour of class time. Tests were conducted in a library computer lab, and a facilitator was on hand to read instructions and answer questions, but not observe participants. A total of four tests had to be thrown out, usually due to a failure to return to the new site's home page between each task.

The results were “at once revealing and frustrating” for the first test group. The tasks with the worst rate of completion included finding magazine and journal articles (not a single participant found the link, which was under the term “databases,” but rather followed a link labeled “E-journals”) and navigating to the university homepage which used a logo link standard across the whole school. More participants failed than succeeded for 60 percent of the exercises. Forty percent found the labels easy to understand and the homepage easy to read in the opinion questionnaire.

Using the results for the first test changes were made to labeling and page structure of the new design and the test run again with the second set of participants. Success rates improved, with the majority of participants failing in only 43 percent of the tasks. Interestingly participants in the second test had trouble with tasks that posed no difficulty to the first test group, even though no page elements in those tasks had been changed. Also interesting was the mismatch between actual performance and user's ratings of the interface, due to either reluctance to give a bad rating or lack of knowledge of how much difficulty they had compared to optimal navigation paths.

Many of the final results of the test were not actually implemented in the site, as the library faculty disliked the idea of using plain language instead of library jargon, preferring to educate patrons instead. Unfortunately, the author notes, new students enroll every year and many access the site completely

remotely—never coming into contact with library instructors.

In “How do I find an article? Insights from a Web usability study” Cockrell and Jayne (2002) introduce their concern that library patrons are confining their search for resources to the web and online catalogs, losing personal contact with library staff and appropriate instruction. They intend to determine if these “surrogate library-instructor” interfaces are, indeed, effective.

Cockrell and Jayne's literature review begins with Nielsen (1993) and Rubin (1994), mentions Battleson, Booth, and Weintrop (2001), and continues through Spool (1997) to lay the ground rules for formal usability studies, including giving users task-oriented exercises, asking users to think aloud, the efficacy of a small number of participants (Nielsen and Landauer, 1993), and the user-centered approach—each application is unique and must be built for the clientele that will actually use it. They mention Garlock and Piontek (1996) as early practitioners of usability testing for library sites, and mention a number of usability problems found on library sites by other researchers, including the use of difficult terminology and abbreviations, the assumption that patrons knew how the libraries' were organized, and lack of knowledge of coverage of different indexes and specific fields for searching in each. They found no other studies at that time specifically addressing search for periodicals.

The authors defined their question: “Do difficulties in locating articles result from lack of knowledge of the concepts and processes involved in indexing and retrieving articles or because users are unable to find the resources they seek among the library's Web pages?” (p. 123) Specifically, they tried to determine if success was related to education (undergraduate, graduate, or faculty), if unsuccessful searchers showed similar search patterns, and if successful searchers were discriminating about which indexes they used for each task.

The Western Michigan University library website was tested. Participants included 10 faculty members, 10 graduate students, and 30 undergraduates. This sample size was chosen because undergraduate students at the school come from a wide range of disciplines and include international and non-traditional students. Participants were recruited from the library and other public areas on campus. Participants were given $5.00 McDonald's gift certificates for incentive to participate. Scripts were developed for all aspects of the test and test administrators were also given a form to record participant comments and search paths. The administrators were a group of 20 trained librarians and staff members. The test itself was administered in the naturalistic setting of a public terminal in the library. There were no time limits for task completion.

Three periodical search tasks were taken from a list of 20 general tasks developed by a committee of six librarians and pretested with students and staff. The three included finding a magazine article on affirmative action, a journal article on endangered species, and a newspaper article from the previous week on the Senate race in New York.

The data recorded for each task included the search path, participant remarks, and “other relevant behaviors” including failure to scroll through an entire page, as well ass level of success. Three levels of success were recorded: successful, “somewhat successful” when an article was found but not from the requested source, and unsuccessful.

Analysis of the results found that there was a significant difference among the different source types, and a significant difference in success rate among the different user groups for the journal and newspaper task, with undergraduates faring worst on both. The authors analyzed successful and unsuccessful paths

for each of the tasks in detail. For example, four participants types keywords into the "search library" box, but it searched only internal pages and not catalogs, and the most common reason the magazine search was unsuccessful was because users attempted to use OPAC to find the article.

Interestingly successful searchers did not always have shorter average paths than unsuccessful searchers. Also noted was confusion over the word "journal" from user comments—it is possible that the task used terminology the participants were unfamiliar with. It would be interesting to run a similar test with user-generated tasks, or for students, assignment-like tasks created by professors.

Key observations of search behavior included:

- Few participants took the time to read explanations, descriptions, search hints, or help screens;
- They carried over their Web search habits to searching library databases;
- They were inclined to reach hasty conclusions, for example, that no record existed when an improperly constructed search did not return one;
- Some individuals (especially undergraduates) were inclined to give up very easily;
- Many participants were not selective, choosing the first item in a list of indexes or the first record in a list of citations; and many participants did not scroll down to information that was relevant.

In the end, four recommendations were made: providing more guidance and more options, limiting content and layout elements, using specialized terminology less and more consistently, and publicizing library assets.

While most of the research presented here is from the United States, Ebenezer's "Usability evaluation of an NHS library website" (2003) is interesting because it details testing done in the United Kingdom. This study also stands out because it details the use of a number of different usability testing methodologies, including comparison to other sites, focus groups, survey research, heuristic evaluation, user testing, and card-sorting and label terminology exercises.

The site tested with all these methods was the South London and Maudsley NHS Trust Multidisciplinary Library website, a library for mental health and social services workers. The user population had varying degrees of computer literacy and computer access—at the time of the test, access was not always available within the organization and users were expected to access the site from home.

This paper's review of literature suggests that the authors were aware of common problems with library sites. Articles by Veldof (2000), Marmion (2001), and Gullikson et al. (1999) indicated that sites are often librarian-oriented, and a study by Matylonek (1999) is mentioned that details five possible sources of this: discipline jargon, hierarchical bias, expert proficiency, folk classification, and emphasis on preferred services. They also mention Nielsen and Landauer's (1993) finding that many, small tests are best as well as some criticism (Spool, 2001)

In the first phase, a six NHS library sites sites were evaluated using a design and content checklist designed by the researchers to provide a benchmark and gather ideas. Secondly focus groups of nine users in three groups were given 15 minutes to play with the site and 45 minutes to discuss. The third phase, the formal usability test, involved seven participants given 15 tasks to complete. Metrics for this part of the study included percentage completed, number of false starts, longest time for each task, number of tester prompts needed per task, and user satisfaction ratings. Also part of this phase was a card-sorting exercise in which users sorted site menu items into categories with results analyzed with

cluster analysis software. The fourth phase was a detailed questionnaire with screenshots of the site's menus and questions about what respondents would expect to find under each link.

A total of 32 participants were recruited via an organization-wide email and personal contacts, and although they represented a wide range of users, a few groups were not contacted. Many participants were medical or professional non-clinical staff, and the majority were female. A free lunch was offered as incentive for participation.

The results of the survey of other site revealed a wide range of content, organization, and resources available from similar sites. The focus groups revealed a number of features that users liked (for example, use of language and links to web logs) and some they did not, most often a lack of a specific resource (for example full online catalog access), although the author admits the groups may have been too small and that the designer should not have lead the groups.

The formal usability test revealed a learning effect, as users were more successful the more they used the site. Participants tended to fall into two groups—searchers and menu browsers. Overall they were very successful, with 92 percent of tasks completed without helpful prompts. Users rated the site highly in comprehensibility of terminology, ease of use, and likelihood of return. Information gathered from card sorting indicated a strong association between 'interlibrary loans' and other libraries' and 'our holdings' and that 'our holdings' was difficult for users to classify.

The site fared worst in the questionnaire, with users having a hard time guessing what would fall under labels like ‘general information,’ ‘facilities for readers,' and the 'ATHENS' brand name for a specific authentication system. In general, problems fell in two main areas—organization and terminology. The author felt the tendency of participants to reveal the same usability issues justified the small sample sizes.

In “What words and where? Applying usability testing techniques to name a new live reference service,” Duncan and Fichter (2004) attempted to employ a user-centered design process to introduce a “live reference” service at the University of Saskatchewan's Health Sciences Library. Unlike many of the other studies in this review, the methods used were not limited to testing existing systems.

The researchers were primarily interested in choosing a label and link placement for the new service so that site users would find and properly utilize the new service.

This article's literature review was limited to one other study on a very similar situation (a 2000 by the National Cancer Institute reporting on adding their own live reference service to their site) and a discussion of the need to avoid library jargon and acronyms (articles by Moyo and Robinson (2001) and Spivey (2000)) and research on what makes users likely to click on a link (articles by Donatello (2001) and Honan). Nielsen's research (1998) is referenced to justify the small user groups used.

In the user-centered design phase of the project, 20 participants were recruited from within the library, including faculty/staff, medical students, and nursing students. The test was done in two rounds: in the first, participants were asked to rank a list of six “invitation to chat” and “invitation to click” phrases for the logo and link. Since none were clear winners, the top phrases were arranged into twelve combinations for the second test, and a winner was chosen from the ones preferred by participants.

The second phase was very similar to many of the tests in other articles. A sample of five users (two responded to a mass email invitation, and three were specifically invited), included a nursing faculty

member, a medical student, a physiology professor, a nursing department staff member, and a nursing student. Participants were able to pick something from "an assortment of treats" as incentive to participate.

Links to the live reference service were placed through the library web site, most often in the upper right-hand corner, although that space was unavailable on some pages and it was places elsewhere. Participants were asked to complete a set of tasks drafted by three members of the library staff familiar with both the site and questions commonly asked by library patrons. The tasks did not specify the use of the live reference service, but it was hoped the service would be discovered and used during the tasks. Participants were asked to think aloud, and both the navigation path and verbal comments were recorded. Another measure included a post-test questionnaire about user satisfaction. Tasks were limited to five minutes.

The study found that users scanned pages for words rather than reading the site; that users rarely looked at the "About Us" section, that they exhibited banner blindness, or the tendency to ignore image links above the content area; that they were not familiar with library jargon or abbreviations, had many problems with the "databases" and "electronic journals" sections, and that searching was a problem even for users who had attended instruction sessions. Four of the users noticed the new link and three used it at least once.

The researchers concluded that the user-centered design process was helpful, as confirmed by the usability test.

"Evaluative Study of Catalog Department Web Pages," by Mundle, Zhao and Bangalore (2004) is interesting because unlike many of the other studies summarized here, it does not deal with public-facing websites or catalog search interfaces. Instead the authors look at an internal library cataloging department's web site. Also, this paper is an example of developing a heuristic analysis for usability rather than user testing employed in many of the other studies. Their goal in this paper is to develop and apply a framework to use in evaluating catalog sites.

The authors state that they found few articles specifically addressing web sites of library catalog departments, although they mention a survey of such web sites by Chressanthis and Wesley(2001) and a description of the construction of one by Harizan and Khoon (1998). Their literature review mentions a number of studies of web usability in academic and library sites, but concentrates on studies that emphasize regular updates and maintenance and process over technical issues in web site success (Nielsen, 1997) (Garlock and Piontek, 1996) (Mach and Kutzik, 2001). Cockrell and Anderson's 2002 study, also reviewed in this paper, is mentioned for their emphasis on user-centered design.

This study took it's criteria for evaluation from studies by Smith (1997), Wyman et al. (1997), Clausen (1999), and Rettig and LaGuardia (1999) and "both public and private workability perspectives" from a study by McClure (1997). Specifically, they looked at four categories:

- Accessibility (e.g., "Can outside users access the Web page?" "Is the page linked to library's Web page?"),
- Design and structure (e.g., "Are the graphics adding any relevancy to the page?" "Does it have a searching capability?"),
- Internal documentation (e.g., "Does it include ... Contact information?" "Does it have training/trainer's manual online?"), and

- External resources (e.g., “Are there ... Any reference tools listed?” “Are there ... Links to electronic discussion lists?”).

The sample for this study included internal library web sites from 10 schools in the Committee on Institutional Cooperation (CIC). Each was rated a yes, no, or “somewhat” for each of the questions in the four categories by three investigators and then discrepancies were smoothed by laster discussion. A performance index was constructed from the average scores for the four categories. The authors explain that usability touches on all four of these categories, not just ease of navigation. In effect, the authors sought to find measures and an objective index “that denotes effectiveness, efficiency, and satisfaction of the intended users or audience,” without directly measuring any of the above with actual users.

Accessibility was found to be generally good. Most pages had uniform design and structure, while 90 percent lacked graphics, relevant or otherwise, and 70 percent lacked search functionality. Under the internal documentation measure, 90 percent had links to various cataloging tools and 80 had links to policies and procedures, and while some contact information was prevalent, mailing address and department newsletters were not often seen past the first page. In general, less than half the sites provided links to the different types of external resources.

Across all sites studied, differences in internal documentation and external resource measures where significantly correlated and accessibility was significantly higher than the other scores. Internal documentation and external resources were seen to be two parts of what other studies might call the sites' “content,” and therefor logically related. The scores of the index as a whole were found to differ significantly for the different institutions. In general, most sites were near the median score with three performing especially well and one poorly. The paper also goes into a detailed discussion of a few of the sites.

In “Metasearching in Boston College Libraries – a case study of user reactions,” Talent (2004) details an informal usability test done on a meta-searching system while under evaluation. This work was prompted originally by focus groups that expressed a string desire to search multiple databases simultaneously. The desire to fill a user need as well as knowledge the users were unaware of the specifics of the 300 databases available prompted the library to try out a product called MetaLib (renamed MetaQuest).

Unlike many of the other papers in this review, this study was very informal and concentrated more on analysis than testing. Their literature review included Nielsen's rational for testing with small numbers of users (Nielsen, 2000) and brief mentions of general usability articles and books. The test itself was a combination of conversation and observation with six students, four undergraduate and two graduate, from various departments. Other methodologies used included search log analysis.

The first major finding way that users preferred keyword searching, sometimes with Boolean logic, and ignored features that could used to refine a search, preferring to hit the back button and start the search over again. They expected full text to always be available, users preferred a simple, clean interface, and were quick to satisfice. Users were unlikely to read site updates or instructions before using a resource, and would only do so after having difficulty. They were pragmatic, attempting to get the most out of limited time rather than experiment. Users were also unaware of all the resources available. Direct observation of users found some parts of the MetaQuest interface awkward, specifically how personalization came before searching—users did not know which databases might be useful to them before they had used any of them. They could not judge databases by their cryptic names and wanted to

begin searching immediately.

In “Model for Presenting Resources in Scholar's Portal,” Feeney and Newby (2005) are primarily interested in describing their methodology for selecting resources for and implementing a federated search (meta search) system at the University of Arizona libraries. The article is included in this review because usability testing was an important part of the process and their methods and results are reported in detail.

The metasearch system described here is similar to the one described in the Tallent (2004) article above, but the usability testing methodology used was more formal. In addition this system included subject area profiles such as “Psychology/Sociology” for users to choose to help narrow down the databases searched.

The detailed explanation of the development of the Scholar's portal is beyond the scope of this article, except for the fact that team members did try to keep users in mind and specifically tried to compile and consider the different user groups (although no actual users were directly involved in the design process). The availability of full text was made a priority because it was a known user desire, for example.

Researchers consulted with fellow faculty members and the literature in creating the testing methodology. Referenced works included Dickstein and Mills (2000), Veldof, Prasse, and Mills (1999), *Usability Testing for Library Websites: A Hands-On Guide* by Norlin and Winters (2002), *Handbook of Usability Testing* by Rubin (1994), and *Don't Make Me Think! A Common Sense Approach to Web Usability* by Krug (2000).

First the team determined the questions they wanted to answer through testing:

- Is a forced log-in a barrier?
- Do users understand the profile labels and what a profile contains?
- Do users understand what is represented in the results list and where the search results come from?
- Are users satisfied with results they get?
- Are users satisfied with a Google-like interface, or do they want an advanced search feature?
- Will they use Scholar’s Portal again? (p. 202)

Assessment was done in three phases: first, feedback from the library staff; second, interviews with users about labels used in the interface, and third, formal usability testing and a satisfaction survey. Only five users, gathered from a central public computer lab in the library, were interviewed in the second phase (with Nielsen cited as rationale for this small number). The third phase involved 18 University of Arizona faculty and students recruited by fliers in public areas on campus and a $10 gift certificate to campus cafes as an incentive. The majority were graduate students with just one undergraduate, but participants came from a number of different departments. Nine tasks were assigned and participants were asked to think aloud while two testers recorded.

The second phase found that same labels for fields of study were more clear than others, and that none of the participants really knew what “collections” meant. Interestingly participants suggested alternative terms with the word “databases” in it—“database” is cited as a confusing word in other studies in this review. The formal review found that large number of users did not pick the subject area profiles testers thought they would for their tasks. Users also tended to use the Boolean “and” operator in their searches,

although the system did not support the feature. Labels such as “open” for a link to the full text of an article were confusing to users that were used to labels like “HTML,” “PDF,” of “full text” from other search systems. Results from the post-test survey were mixed, with some confirmation of the results of the interviews and tests. User satisfaction ratings showed there was much room for improvement, but that users were willing to try the system again.

Conclusion

In general, library web sites usability studies are found to be very effective with relatively little cost, but very similar usability issues come up in almost all of the studies. A number of different methodologies are used but formal user testing seems most common, with cues often taken from a small number of usability experts and their books and research. No matter the methodology employed, it is striking how often the exact same problems come up in nearly all of the studies. There seems to be a large disconnect between how librarians, faculty, staff members, and library site designers think about searching and resources and how users do. Study after study discovers that library terminology is confusing the users and that site organizational structure that seem perfectly logical are baffling to users.

Methodology

Although this is by no means an exhaustive or statistical study of all usability research being done in library settings, it is clear from the number of studies and ease of finding recent research that this is a hot topic. No doubt many usability tests and evaluations occur that are not documented and submitted to academic journals. For example, the lack of examples of expert review may be due to the fact that this is often a commercial service.

A number of the articles also mentioned a lack of published literature in usability research in library web sites. It seems many researchers have noticed a paucity in this area around the time these articles were written, since now it is quite easy to find a large number of articles on this topic.

It is encouraging that a number of the tests described in these papers included a trial run of the test in order to test the methodology. This can prove very useful, ironing out any procedural problems before participants show up and test data is ruined. Some of the studies integrated a number of different approaches or did testing in a series of phases and seemed to benefit from the strengths of the different techniques. Also, it is interesting to see testing done both as part of a user-centered design process and as a post-implementation assessment tool.

Most performed the tests themselves in a more or less naturalistic environment, usually in computer labs or classrooms that the users would be comfortable in, rather than a lab environment.

One possible issue with some of the studies is, with the disconnect between library site designers and patrons quite obvious, how did the researchers build their task lists? All of the usability tests tried to represent tasks users would actually want to carry out, but only a few asked users directly or involved them in the process.

Every article describing a user test except for Allen mentioned the five-user rule, and the majority relied upon it for their methodology. Only one article (Ebenezer) mentioned any criticism of the five-user, less-is more approach advocated by Nielsen and Landauer. Although a number of the authors reported that their experience seemed to conform to this finding, with additional users finding many of the same flaws over and over again, it is amazing how pervasive this method is. Since there is still some debate over the validity of this rule, library researchers studying usability should take some caution and not rely blindly on one article's finding for their studies' validity.

The majority of these articles cover projects in academic libraries. It is hard to tell from the small sample, but it seems possible that there are interesting issues in public or corporate library web site usability that are not being studied in the same depth as in university settings. Public libraries may face

even more difficulties because the range of computer and information seeking skills is much wider. Many of the articles talked about the differences in skill and experience within their populations (and rightly so), but even freshman are guaranteed to have a high school diploma and information needs driven by regular assignments. How would the results of a usability test differ when the participants included children, older adults with no computer experience, and casual browsers with no particular information need in mind?

Findings

One finding that stands out—users do not use the same terminology as librarians or library site designers. Although the academic journal articles my be contained in a number of databases, users apparently do not think the word "databases" when they want a journal article. Further, users often look for a single search interface of all resources, or pick the first resource they see to search—in some studies, users tried using the library site search to search databases as well. The sheer number of times this finding was 'discovered' by usability tests could lead to a number of conclusions – first, it does not seem library site designers are familiar with academic research in this area. It is also quite possible that web site project requirements have not included usability. Web sites are often little more than extensive user interfaces, and it is very unfortunate when user needs are not important to a user interface.

Along with labeling issues, another common theme is site organization. Users often seem to have a hard time guessing which resources are under which category and often find paths to resources that do not fit the "ideal" or "optimal" paths preferred by designers and testers. Library web sites, like many others, are often organized as hierarchies branching out from the home page—it is possible that this is not a good model for how users see the relationships between resources. For example, say there is a particular full-text database of journal articles on artificial intelligence. Should it be under the branch for full-text databases in general, or the branch for computer science resources, or the branch for cognitive psychology resources, or somehow linked with results of a search for "artificial intelligence" in the library's catalog? There is no real reason it should only be found in just one of these places, and user-centered design can help determine the most appropriate (and probably multiple) places to place the link.

Library faculty and staff have a different way of thinking about library resources than users. This is logical. Faculty and staff spend much of their time organizing resources, while users do not. The former have time to become intimately familiar with different resources, the latter only use the library as a means to an end—answering a question or researching a topic.

This brings up an interesting question for further research: if there is a disconnect between librarian and patron concepts, language, and expectations in online libraries, is the same true in physical libraries? Although it may be impossible to measure the number of clicks from server logs in physical browsing, similar issues like labeling and proximity of related resources may be worth examining. Despite the other flaws found, not a single web site in this study had done anything as user-unfriendly as labeling row after row of bookshelves with nothing but cryptic call numbers, a common practice in physical libraries.

In their other findings, the studies did not always agree. The success rates varied form system to system, which is perfectly reasonable given the different goals, user groups, and software used. Some of the studies differed as to whether or not there's a learning effect or if users continue to use the same search techniques whether or not they are successful.

Future study

The articles surveyed here were unanimous in their positive evaluation of the usefulness and cost effectiveness of usability testing for library web sites. Judging by these testers' experiences, it pays to plan ahead. In implementing future tests, these questions should be asked:

- What are the goals of the site? Of the test?
- What methods or techniques should be used?
- What is the population, major user groups, and how should sampling be done?
- How will user goals and tasks be determined?
- Have procedural issues been thought out, such as the use of scripts, incentives, the testing environment, user instructions ('think aloud,' etc.)?
- What data will be recorded, and what are the measures of success?
- Will tests be pretested?

The researchers used various measures for their tests. Based on the goals of the test, the test design, and the amount of time and staff available all of the following measures can give an indication of web site usability:

- Success/failure of users at each task;
- Post-test questionnaires with subjective, Likert-type scales about ease of use, confidence, likelihood of return;
- Time for task completion;
- Length of navigation path;
- Minimum possible or optimal navigation path;
- Number of misses (dead end paths);
- In-test qualitative data like “thinking aloud” and body language;
- Use of non-page navigation features (back button, etc.);
- Measures of test participant experience, expertise, or familiarity with the site;
- Measures of learnability, for example decreased task time for a participant as test proceeds;
- Fidelity with user-constructed labels and organization (from card-sorting, etc.); and
- Server log analysis of search terms, navigation paths, etc. used by current users.

Both the consistent disconnect between library web site designers and users and the reported effectiveness of testing shows that usability testing clearly has a place in the field. Although test methodologies vary considerably, the majority of tests rely upon a small number of key researchers and writers for their research design and report similar levels of satisfaction with outcomes. It is in some ways surprising that researchers in a field that has studied information organization, information use and services, and information seeking for decades can learn so much from relatively simple, small-scale tests. The World Wide Web is not going away any time soon, and users expect to do more and more of their information search and retrieval online. Given the examples here, the time for usability testing in library web sites and other user interfaces is clearly now.

Articles Reviewed

Allen, Maryellen (2002), "A case study of the usability testing of the University of South Florida's virtual library interface design," *Online Information Review,* 26, 1, 40-53.

McMullen, Susan (2001), "Usability testing in a library Web site redesign project," *Reference Services Review,* 29, 1, 7-22.

Battleson, Brenda, Austin Booth, and Jane Weintrop (2001), "Usability testing of an academic library Web site: a case study," *The Journal of Academic Librarianship,* 27, 3, May, 188-198.

Ebenezer, Catherine (2003), "Usability evaluation of an NHS library website," *Health Libraries Review,* 20, 3, September, 134-142.

Tallent, Ed (2004), "Metasearching in Boston College Libraries – a case study of user reactions," *New Library World,* 105, 1-2, 69-75.

Feeney, Mary and Jill Newby (2005), "Model for Presenting Resources in Scholar's Portal," *Libraries and the Academy,* 5, 2, April 07, 195-211.

Roca, Joan and Roland Nord (2001), "Usability study of the MnLINK Gateway," *OCLC Systems & Services,* 17, 1, 26-34.

Mundle, Kavita, Nirmala S. Bangalore, and Lisa Zhao (2004), "Evaluative Study of Catalog Department Web Pages," *Library Resources & Technical Services*, 48, 1, 48-58.

Duncan, Vicky and Darlene M. Fichter (2004), "What words and where? Applying usability testing techniques to name a new live reference service," *J Med Libr Assoc,* 92, 2, April, 218–225.

Cockrell, B. J., et. al. (2002), "How do I find an article? Insights from a Web usability study [at Western Michigan University]," *The Journal of Academic Librarianship,* 28, 3, May, 122-32 .

References

Chisman, Janet, Karen Diller, and Sharon Walbridge (November 1999), "Usability Testing: A Case Study," *College & Research Libraries*, 60, 552–569.

Chressanthis, June and Wesley, Kathryn (2001), "The Technical Services Departmental Webpage: A New Wave Management Tool for Serialists," *Serials Librarian,* 40, 3/4, 194.

Clausen, Helge (1999), "Evaluation of Library Web Sites: The Danish Case," *The Electronic Library*, 17, 2, 83–87;

Dickstein, Ruth and Mills, Victoria (2000), "Usability Testing at the University of Arizona Library: How to Let the Users in on the Design," *Information Technology and Libraries* 19, 3, 144–51;

Donatello, M. (2001), "How do I click thee? let me count the ways," Accessed 13 Jul 2005, http://www.naa.org/artpage.cfm?AID=1585&SID=1020.

Dumas, Joseph C. and Redish, Janice C. (1993), *A Practical Guide to Usability Testing*, Ablex Publishing Co., Norwood, NJ.

Dumas, Joseph C. and Redish, Janice C. (1999), *A Practical Guide to Usability Testing*, Revised Edition, Intellect Books, Portland, OR.

Eliasen, Karen, McKinstry, Jill, and Fraser, Beth Mabel (1997), "Navigating Online Menus: A Quantitative Experiment," *College & Research Libraries*, 58, 509–516.

Garlock, Kristen L. and Piontek, Sherry (1996), *Building the Service based Library Web Site: A Step-by-Step Guide to Design and Options,* ALA, Chicago.

Gullikson, S., Blades, R., Bragdon, M., McKibbon, S., Sparling, M. and Toms, E. G. (1999), "The impact of information architecture on academic web site usability." *Electronic Library*, 17, 293–30.

Hackos, J. and Redish J. (1997), *User and Task Analysis for Interface Design*, John Wiley & Sons, New York, NY.

Harizan, Amanda and Khoon, Low Ke (1998), "Technical Service and the World Wide Web: Building the NTU Library Cataloging Home Page," *The Electronic Library,* 16, 6, 387–93.

Head, Alison (1999), *Design Wise: A Guide to Evaluating the Interface Design of Information Resources*, Cyber-Age Books, Medford, NJ.

Hom, James (1999), "The Usability Methods Toolbox," (Online) http://jthom.best.vwh.net/usability/.

Honan, M., "Will microads save online content? (part one)," *Online Journalism Review,* Accessed 5 Feb 2003, http://www.ojr.org/ojr/business/1015015782.php.

Krug, Steve (2000), *Don't Make Me Think! A Common Sense Approach to Web Usability,* Que, Indianapolis, IN, 155–73.

Mach, Michelle and Kutzik, Jennifer S. (2001), "The Web Is Everyone's Business: A Distributed System for Managing the Library Web Site," *Colorado Libraries,* 27, 3, 29.

Marchionini, G. (1995), *Information Seeking in Electronic Environments*, Cambridge University Press, London.

Marmion, D. (2000), "Library web page design: are we doing it right?" *Information Technology and Libraries*, 20, 2–3.

Matylonek, J. (1999), "Cluster analysis in Web design." Presentation given at the ASIS PNC Bridging the Gap: Innovations in Information Services Conference, September 17–18, (Online) http://osulibrary.orst.edu/staff/matylonj/cluster/sld001.htm.

McClure, Charles R., Wyman, Steven K., and Beachboard, John C. (1997), "Quality Criteria for Evaluating Information Resources and Services Available from Federal Web Sites," *[OCLC] Annual Review of Research,* Accessed Oct. 10, 2003, http://digitalarchive.oclc.org/da/ViewObject.jsp?objid=0000003456.

Moyo L and Robinson, A (2001), "Library jargon as a factor in information design for Web usability: survey report [summary]." *16th Annual Computers in Libraries 2001: collected presentations*, 2001, Washington Hilton & Towers, March 14–16, 2001. Medford, NJ, *Information Today*, 157–65.

Mundle, Kavita, Bangalore, Nirmala S. and Zhao, Lisa (2004), "Evaluative Study of Catalog Department Web Pages," *Library Resources & Technical Services,* 48, 1, 48-58.

Nielsen, Jakob (1993), *Usability Engineering,* Academic Press, Boston.

Nielsen, Jakob (1997), "Top Ten Mistakes of Web Management," *Jakob Nielsen's Alertbox (June 15, 1997)*. Accessed Oct. 10, 2003, http://www.useit.com/alertbox/9706b.html.

Nielsen, Jakob (1998). "Cost of user testing a Website." *Jakob Nielsen's Alertbox for 3 May 1998*. Accessed 5 Jul 2003, http://www.useit.com/alertbox/980503.html.

Nielsen, Jakob (1999), *Designing Web Usability: The Practice of Simplicity,* New Riders Publishing, Indianapolis, IN.

Nielsen, Jakob (2000), "Why you only need to test with five users", *Jakob Nielsen's Alertbox for 9 March 2000*, (Online) http://www.useit.com/alertbox/20000319.html.

Nielsen, Jakob and Landauer, T. K. (1993), "A mathematical model of the finding of usability problems." Proceedings of ACM INTERCHI '93 Conference, Amsterdam, the Netherlands, 24–29, April, 206–13. Cited in Nielsen, J. (2000). "Why you only need to test with five users." (Online) http://useit.com.alertbox/20000319.html.

Norlin, Elaina and Winters, C.M. (2002), *Usability Testing for Library Websites: A Hands-On Guide,* American Library Association, Chicago.

Office of Communications, Communications Technology Branch, National Cancer Institute (2000), "Usability.gov: lessons learned: instant messaging," Bethesda, MD: The Institute. Accessed 26 Apr 2002,

http://www.usability.gov/lessons/IMplearned.html.

Park, I.K. (1997), "A comparative study of major OPACs in selected academic libraries for developing countries user study and subjective user evaluation", *International Information & Library Review*, Vol. 29, pp. 67-83.

Peters, T.A. (1998), "Remotely familiar: using computerized monitoring to study remote use'", *Library Trends*, Vol. 47 No. 1, pp. 7-20.

Preece, J. et al. (1994), *Human-Computer Interaction*, Addison-Wesley, Harlow.

Rettig, James and LaGuardia, Cheryl (1999), "Beyond 'Beyond Cool'," *Online,* 23, 4, 51–55.

Rubin, Jeffrey (1994), *The Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*, Wiley, New York.

Rudd, J., Stern, R. and Isensee, S. (1996), "Low vs. high fidelity prototyping debate"', *Interactions*, Vol. 3 No. 1, pp. 76-85.

Smith, Alastair G. (1997), "Testing the Surf: Criteria for Evaluating Internet Information Resources," *The Public-Access Computer Systems Review,* 8, 3, 5–23.

Spivey, M. A. (2000), "The vocabulary of library home pages: an influence on diverse and remote end-users," *Inform Technol Libr,* Sep, 19, 3,151–6.

Spool, Jared (1997), *Web Site Usability: A Designers' Guide,* User Interface Engineering, North Andover, MA.

Spool, Jared., and Schroeder, W. (2001), "Testing web sites: Five users is nowhere near enough." In *CHI 2001 Extended Abstracts,* 285- 286, ACM Press, New York.

Tolle, J.E. (1983), *Current Utilization of Online Catalogs: Transaction Log Analysis*, OCLC Online Computer Library Center, Office of Research, Dublin, OH.

Veldof, J. R. (2000), "Building user-centered library web sites on a shoestring," Presentation given at WILS World, (Online) http://www.wils.wisc.edu/events/ww2k/ww2kpres/veldof.

Veldof, Jerilyn, Prasse, Michael J., and Mills, Victoria A. (1999), "Chauffeured by the User: Usability in the Electronic Library," *Journal of Library Administration* 26, 3/4, 115–40;

Wyman, Steven K. et al. (1997), "Developing System-Based and User-Based Criteria for Assessing Federal Websites," *Proceedings of the ASIS Annual Meeting,* 34, 78–88.