
A case study of the usability testing of the University of South Florida's virtual library interface design

Maryellen Allen

The author

Maryellen Allen is a Reference/Government Documents Librarian at the University of South Florida's Tampa Campus Library, Tampa, Florida, USA.

Keywords

Interfaces, Virtual libraries, Design

Abstract

Usability testing of Web interfaces for virtual libraries is a crucial factor in the continuing development and improvement of the user interface. In 1999, the University of South Florida Libraries decided to embark on a usability study to coincide with the rollout of a new interface design. Because this type of study had not been conducted with the initial interface, its implementation and completion were paramount in the development of the new design. Details the preliminary activities, testing methodologies and results of usability testing by the USF virtual library project's usability study group.

Electronic access

The research register for this journal is available at <http://www.emeraldinsight.com/researchregisters>

The current issue and full text archive of this journal is available at <http://www.emeraldinsight.com/1468-4527.htm>

Introduction

Usability testing of library Web interfaces is a trend quickly becoming the rule rather than the exception. As the amount of resources available on the Web has grown to an astonishing number, libraries and their users are quickly finding themselves drowning in the magnitude of their own sites. One way to manage the flood of online information resources while maximising user access is to conduct usability testing.

Usability testing has been a sustained practice in the realm of software development for well over a decade, but only recently has the electronic information industry discovered its merits. Perhaps the study most responsible for bringing usability testing to the forefront of the profession is Dickstein and Mills' (2000) report of the University of Arizona library's usability test in 1998. Not only did this study serve as an excellent model for others wishing to conduct usability testing in their own institution, but it also pointed out some of the most crucial factors in interface design, namely:

- library interfaces are designed for librarians rather than users;
- testing provides a means of settling disputes about the interface design;
- testing arms the designers with objectivity regarding the interface;
- testing sends a message to the user base that the institution is interested in their input.

Other studies within the body of the literature on usability testing also proved extremely helpful in the development of the University of South Florida (USF) study. Most notably, Rubin's *Handbook of Usability Testing* (1994) and Dumas and Redish's *A Practical Guide to Usability Testing* (1993) provided invaluable ideas and input in the process of instrument development. Rubin's (1994) outline of the four types of usability tests (i.e. exploratory; assessment; validation/verification; and comparison) was indeed an asset during preliminary development activities, as the proper identification of test type clarified both the purpose of the testing and the development of the testing instrument. For the purposes of the USF study, a combination of the assessment and comparison tests were

Refereed article received 18 October 2001

Approved for publication 15 November 2001



deemed most relevant. The assessment stressed quantitative measures, while the comparison test embodied the objective of improving on the present interface by testing a series of prototypes in a side-by-side comparison with the old. Dumas and Redish gave important insights into the interpersonal aspects of the testing process, such as the consideration given to human-computer interaction and the impact of certain behaviours on those giving and taking the tests. Peripherally, the body of literature concerned with usability testing in general served to round out the review and provide a balance to the overall process. Head (1999) and Fichter (2000) each furnished a good synopsis of usability testing and delineated the possible hazards test givers and test takers are likely to encounter. Fraser *et al.* (1998) further supplement the literature by providing an important reminder about accessibility in the interface design with respect to users with disabilities, a feature curiously absent in many studies.

Background

When the first virtual library interface was conceived in 1995, there was very little in the way of formal usability testing conducted. The original interface was the product of a development-team approach charged with creating a single access point for patrons wishing to use the library's ever-increasing store of electronic resources (see Figure 1). Heuristic evaluation, i.e. a comprehensive walk-through of all of the links on the new site, was the primary means of development and evaluation of the interface during its

development. On implementation of the primary interface, the designers had only to contend with providing access to 20 databases. As the number of resources available via the Internet increased, the interface became increasingly cumbersome to maintain.

Additionally, comments from the user base indicated that the interface was unintuitive and somewhat difficult to navigate, especially the button to access the library's online catalogue. Originally, this button was labelled simply by its proprietary name, WebLUIIS. It became apparent rather quickly that almost no one knew what that term meant.

Because of the problems with upkeep in tandem with user feedback indicating the navigational difficulties with the interface, the virtual library interface design project group (IDPG) decided to embark on a project to design a new interface. The design would be driven by results obtained from usability testing. Early in 1999, a subgroup (appropriately named the usability study group or USG) was formed to undertake the task of designing, testing, and implementing a usability study for a new prototype interface in development at that time.

Research problem

The USF, located on the central West coast of Florida (see Figure 2) is a large, multi-campus system serving a student population of over 37,000, with a faculty and staff count of over 5,000 members (<http://usfweb.usf.edu/ataglance.html>). The university also serves a sizeable population of distance-learning students and maintains three satellite centres in addition to four regional campuses.

Figure 1 First virtual library interface

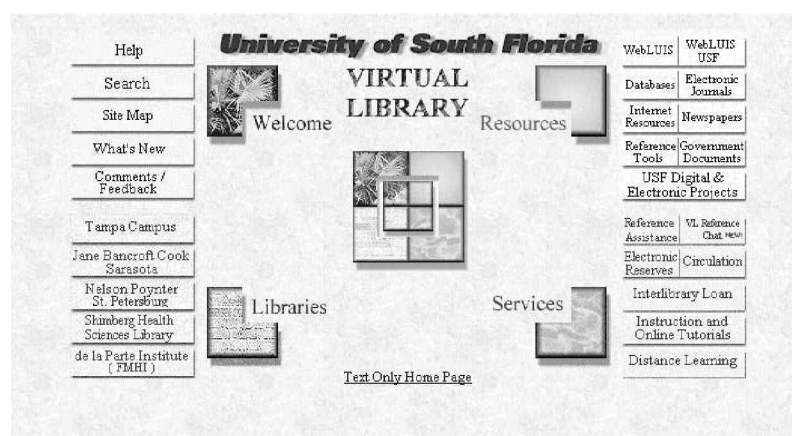
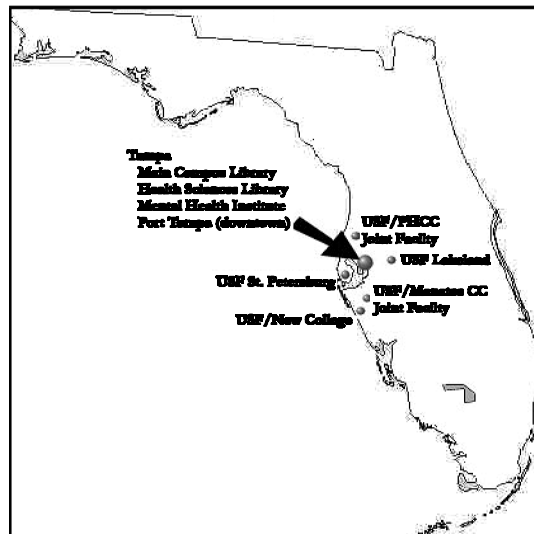


Figure 2 USF coverage area

Traditionally, the student population of the university has been, and continues to be, comprised of commuter students, many of whom are of non-traditional age returning to complete their degrees. These factors alone, i.e. the sheer quantity and quality of students and faculty combined with the wide geographic distribution of resources, provided an environment in which a virtual library, accessible to all students and faculty regardless of location, was a necessity.

The original virtual library interface design, unveiled in November 1996, was the culmination of a great deal of planning and development. At the same time, however, this crowning achievement contained some very serious flaws. First, the primary interface was essentially a large image map – entirely graphical in nature with a series of clickable buttons designed to provide access to the virtual library's resources. The use of the image map for a user interface proved quite inflexible, difficult to alter and unintuitive for the users. This design, extremely static in nature, required a labour-intensive process utilising graphics software to either add button images to the design, and/or shrink the image to accommodate the addition of links. Then, the image would have to be re-mapped and re-mounted on the server. When the first interface debuted, these concerns were not as significant because the number of available online resources was limited to the library's online catalogue and approximately 20 article databases. Several years later, after the

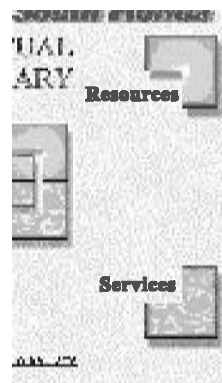
number of databases had climbed to well over 300 and new features such as electronic journals, electronic newspapers, electronic book and image collections were added, these same concerns became paramount as the libraries continued to try to fit an increasingly complex and quickly growing adolescent portal into an infant framework. Staff/patron interaction strongly suggested that the original complaint that interface was not intuitive was being exacerbated by this rapid growth. Additionally, parts of the graphical interface were linked to pages providing additional resources (see Figure 3), yet those areas were not obvious and went largely unused, save by the librarians who were familiar with the intricacies of the design. Recognising the shortcomings and increasingly cumbersome nature of the original interface design, the virtual library project manager prompted the IDPG to begin work on a new design and started the momentum for formal usability testing.

Methodology

Preliminary development

Prior to usability testing, the IDPG worked extensively to develop several new interface designs. The original concept and guidelines for development of the interface were still relevant and the group was instructed to adhere to them. These guidelines placed restrictions on the use of the following Web design elements:

- use of Java or JavaScript;
- drop-down boxes;
- frames; and
- animated or blinking graphics.

Figure 3 Additional resource links

Furthermore, the additional guidelines were to be observed:

- the design will feature text over graphics wherever possible;
- a link to each individual USF library will be included;
- the university logo will appear on the upper left side of the page and will link to the main university homepage;
- a text-only link will be provided;
- the contents of the front page will be contained within a standard 800 × 600 pixel browser window and will not require the user to scroll up or down, left or right.

Finally, as the result of server log analysis, virtual library project designers suggested that the new interface provide the following links to these additional pages from the virtual library homepage:

- the online catalogue;
- the Florida State University System union catalogue;
- article databases, electronic journals, electronic newspapers and electronic book collections;
- help screens;
- my virtual library customisable interface service;
- services such as online renewal, interlibrary loan and electronic reserves;
- the Government Documents departmental page;
- all resources by subject;
- development information on the history of the virtual library project;
- university-wide pages of interest.

The number of restrictions and required links, as well as the limitation on the size of the page, placed significant limitations on the design possibilities. The guidelines forbidding the use of Java, JavaScript and drop-down boxes were necessary to keep the site accessible to disabled users and users of legacy systems; however, in hindsight, the number of required links combined with the requirement that the page not scroll was a recipe for frustration. Indeed, each member of IDPG worked tirelessly to come up with a new design that met the requirements and guidelines while keeping the usability goal in sight. After several designs were reviewed and subsequently discarded, the group concluded that its objectivity with respect to the design had suffered greatly as a result of the close relationship between the group and the

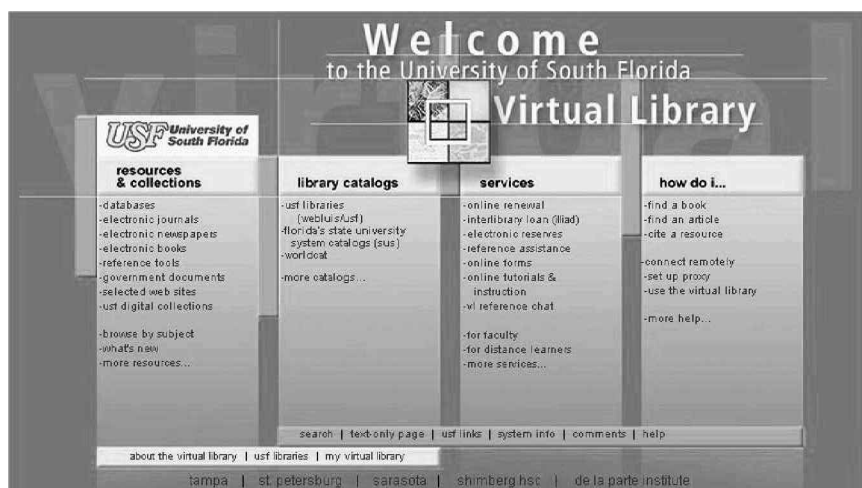
continual development of the interface. In order effectively to design something completely different while fulfilling all the requirements, the design for the new interface was contracted out to a third party for development. After several iterations of designs were reviewed, the group finally settled on one and prepared to test it (see Figure 4).

Experimental testing

Rather than engage in a time-consuming test of the old interface design, the group relied on results from previous focus groups as a starting point prior to development of the testing instrument for the new interface. Before the USG developed and implemented the tests, the group reviewed and consulted a wide variety of books and articles concerning usability testing and test participant psychology. The group was divided between offering an unmediated test, and more stringent testing that involved sitting alongside the test participants while recording their actions. The main obstacle with the latter kind of test was the scarcity of individuals available to serve as test recorders. The library had neither the staff nor the funds to recruit and train new people for this purpose. However, an unmediated testing environment would allow the group to test a large population with the understanding that a fair portion of the results would have to be discarded due to test participant error. A fully mediated test with a test recorder would provide more precise results, but would require staffing resources that were not at the group's disposal. The group knew that it would eventually have to choose between conducting the large-scale, unmediated test utilising a large testing population, or a manually recorded test detailing the actions of an extremely small (and therefore likely skewed) population. The group eventually decided that an experimental test would be conducted first, utilising both the unmediated and mediated environments. The results of that experiment would then determine which testing method would provide productive results while remaining feasible.

The group formulated an experimental testing instrument and made plans to evaluate it. Borrowing structural ideas from a wide array of other testing instruments, the experimental instrument consisted of two parts: The first part contained a series of 15

Figure 4 New interface design



exercises, each instructing the test participant to locate a different resource on the new virtual library interface design using the links from the test page. The second portion consisted of six questions attempting to determine the tester's opinion of the design.

Testing for the experimental instrument utilised a volunteer population of 26 undergraduate students divided into two unevenly distributed groups. The first group, containing 22 participants, was left unmediated and asked to record manually their own navigational clicks to the requested resources. The second set of four participants were matched with a recorder charged with sitting alongside the test participant to record manually the testers' activities. In both cases, the test participants were instructed to navigate to each requested resource by clicking on the link or links that they believed the most likely path to satisfy the question. The self-recording group was instructed to navigate as far as they could within seven clicks to locate the resource. If they were unable to complete the exercise successfully after seven clicks, they were to go on to the next question. For the mediated group, the same limits applied, but it was the responsibility of the recorder to stop the test participant after seven clicks. Each experimental group provided valuable information about both the testing instrument and the methodology.

The format of the instrument proved mostly acceptable, with the wording on a few of the exercises in need of altering to reduce ambiguity or increase clarity of intent for inclusion in the real testing instrument.

Specifically, the first question, "Locate the correct link(s) for finding materials in the library" was changed to "Locate the correct link(s) for opening the USF libraries' online catalogue". Additionally, question ten, "Locate the proper link for placing an interlibrary loan request" was changed to "Locate the proper link(s) to request a book or article from another library". In the first example, the meaning of the question was perceived to be ambiguous, as "finding materials in the library" could mean articles as well as books. Other participants interpreted the question literally and sought out the online floor plan of the library. In the second case, the test administrators found that few users were familiar with the term "interlibrary loan" request, even though they were familiar with the service itself. Changing the wording of the question helped to clarify the meaning of the question for the users. Finally, on the experimental test, one question was eliminated and replaced with a more relevant one. On the preliminary instrument, the last question was "Locate the link(s) that will take you to a list of recently purchased databases". The intent of the question was to test user familiarity with the "What's new?" link on the virtual library that detailed which databases or services have been recently purchased or implemented. However, the group determined from examining traffic logs that this link was not only one of the least visited pages, it also was determined not to be a crucial resource for users seeking information. For the actual testing instrument, this question was replaced by the question, "Locate the link(s) that will provide online

instructions for accessing the virtual library resources remotely". Indeed, given the large number of questions fielded from the library's user base regarding remote access to resources, this question was a much better indicator of effective interface design. During the experimental testing period, little attention was paid to the results of the opinion portion of the test, and in hindsight, this was a crucial error that would come back to haunt later test results.

More substantial information was gleaned from observing the performance of the test participants. The first group, instructed to record manually each link they followed on every exercise, exposed many problems associated with this methodology. The most disappointing of these were the gaps often left in their notations. This became obvious on analysis of the results when two consecutively recorded link paths were a navigational impossibility on the interface as the test administrator attempted to recreate the test participant's activities. Furthermore, some participants skipped over entire questions, or simply wrote "can't find it" on the response sheet, even though the test administrator emphasised the preferred procedure of attempting to locate the resource even if the participant was unclear about where to start.

Difficulty interpreting the writing of the test participants emerged as another significant hurdle. For the most part, handwriting styles were varied, but legible. In a few instances, however, it was extremely difficult to decipher what the test participant had written on the response sheet. Finally, the group of participants that were responsible for recording their own paths required twice the amount of time to complete the test over the mediated group. On average, it took 1.75 hours for this group to complete the test, whereas those in the mediated group required only 45 minutes to an hour. While the mediated testing group did not encounter the same set of impediments as the self-recording group, the unique obstacles inherent in fully mediated testing compelled the test administrator to question if the second group fared any better.

Most of the test participants in the second group were made nervous by the recording process, and either sought approval from the recorder during each exercise or paid an undue amount of attention to what the recorder was writing down. One participant

eagerly watched the face of the test recorder after each mouse click to see if the recorder would provide feedback. Even though the recorder refused to furnish positive or negative enforcement and assured the tester that there were no wrong answers, the behaviour of the test participant was significantly altered by the presence of the recorder and the mere idea of being closely observed. On the whole, these test participants were extremely self-conscious and tended to second-guess their navigational decisions to a great degree. This behaviour was only exacerbated by reminders from the test recorder that the participant should not attempt to backtrack the chosen path, but should continue until each exercise was completed, or the participant reached the seven click maximum put in place by the test administrator.

Another unforeseen problem was that of attempting to record manually each navigational click of the test participant while keeping up with their activities. In order to refrain from altering the behaviour of the participant as much as possible, the recorder did not set a tempo for participant navigation or instruct the test participant to wait for the recorder. Each participant was free to proceed at his/her own pace. This led to some difficulties as the test recorder endeavoured to pay close attention to the activities of the test participant while effectively recording each click at the same rate as the participant.

Confronted with the problems of both self-reported tests and fully mediated tests, the USG identified a need to invent a means of giving the test, producing reliable and analysable results without altering the behaviour of the test participant. The group members mused about the desirability of an automatic recorder hooked to each computer that would track user activities. After some discussion, it occurred to the group that the library did, in fact, have such a tracking device for every machine in the library. The library's Apache Web server effectively recorded all user activity within the library's domain. Network traffic data could be captured from the server and produced in a report log. Among other things, these logs would show the name of the computer requesting the file, the URL of the file requested by the user, and the date and time of the request. The test administrator would simply have to record the IP address of the machine and the start and

stop time of the testing interval. The drawback of the methodology was that all data from network traffic outside of the library's domain would be lost. Therefore, the testing instrument had to be altered somewhat to include navigational exercises that remained within the domain of the virtual library's pages. This was not difficult, as many of the exercises already were set up this way. A few of the exercises involving more than two navigational clicks that ultimately led to a third-party resource were left alone as the preceding path would be enough to extrapolate the success or failure of the final lost click.

The automated testing model

The real testing instrument resembled the experimental one in many ways (see Figure 5). Like the experimental instrument, it consisted of two portions; the first a series of 15 navigational exercises and the second a questionnaire with six questions employing a Likert scale to determine user satisfaction with the visual aspect of the interface design. The differences lay both in the instructional information and the test administration. In order to conduct a successful automated mediated test, the instructions for the first portion of the test had to be clear and concise. Well before testing began, the test administrators ran through the exercises and captured the data to get an idea how long the test would take, as well as how the log report appeared. This preliminary activity also

provided a kind of benchmark by which to judge the navigation behaviours of the test participants. The log reports generated from the experimental run-through had a few serious problems. First, all the network traffic data were reproduced on the log report. This included each GET call for each graphical image on the screen as well as each URL. This made the log report quite confusing and almost useless for analysis. Second, there was no way to differentiate one exercise from another: They ran together on the report. It was only possible for the test administrators to differentiate between each URL and each exercise question on the benchmark report because of the intimate knowledge the administrators possessed regarding the test design. Inexperienced users at whom the test was aimed would likely navigate to unknown locations, perhaps using obscure paths, effectively eliminating this advantage.

The USG therefore decided that the graphics must somehow be eliminated from the final log report and, further, that some kind of indicator was required to alert the test analyst at what point each exercise began. The library's technology department, which formulated a script to eliminate all the graphical hits from the server logs, addressed the first issue rather effectively. The other problem was addressed by the test designers' decision to instruct the test participant to return back to the homepage on the completion of each exercise, but not to return to the homepage until each exercise was

Figure 5 Testing instrument

USABILITY TEST- NEW VL INTERFACE Test Questions																																																									
<p>Part I- User Friendliness/Navigability</p> <p>Instructions: For each exercise, identify the path required to access the requested resource. Navigate by clicking the mouse on the appropriate link(s).</p> <p>*Return to the Virtual Library homepage after completing each exercise.</p> <p>*Do not return to the homepage in the middle of an exercise, even if your path selection does not take you to the intended location. If you are unable to locate the resource, that is still acceptable. Simply return to the homepage after you have navigated as far as you can.</p> <ol style="list-style-type: none"> 1. Locate the correct link(s) for opening the USF Libraries online catalog 2. Locate the link(s) you would click on to research journal or magazine articles 3. Locate the correct link(s) for finding information on how long a book can be checked out. 4. Locate the link(s) that will allow you to search other universities' library catalogs in the State University System. 5. Locate the link(s) that will take you to an online dictionary of the English language. 6. Locate an online help sheet that will tell you how to find an article. 7. Find the link(s) that will take you to the USF St. Petersburg Library's homepage. 8. Find the link(s) that will take you to the main USF homepage. 9. Locate the proper link(s) for launching a search of the USF Virtual Library's Web site to locate a resource. 10. Locate the proper link(s) to request a book or article from another library. 11. Locate the link(s) that will take you to a place where you can search the World W 12. Locate the proper link(s) for renewing library materials online. 13. Locate the link(s) that will take you to information regarding the USF Library orient classes. 14. Locate the proper link(s) to browse Virtual Library resources by subject. 15. Locate the link(s) that will provide online instructions for accessing the Virtual Libr 	<p>Part II- General Impressions</p> <p>Instructions: Please circle the number which best reflects your response to the statement, with 1 being "I strongly agree with the statement" and 7 being "I strongly disagree with the statement".</p> <ol style="list-style-type: none"> 1. I find the new USF Virtual Library homepage visually pleasing. <table border="0"> <tr> <td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td> </tr> <tr> <td>Strongly Agree</td><td></td><td></td><td></td><td></td><td></td><td>Strongly Disagree</td> </tr> </table> 2. I find the new USF Virtual Library homepage easy to read <table border="0"> <tr> <td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td> </tr> <tr> <td>Strongly Agree</td><td></td><td></td><td></td><td></td><td></td><td>Strongly Disagree</td> </tr> </table> 3. I find the category headings easy to understand <table border="0"> <tr> <td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td> </tr> <tr> <td>Strongly Agree</td><td></td><td></td><td></td><td></td><td></td><td>Strongly Disagree</td> </tr> </table> 4. I find the links under each category heading easy to understand and navigate <table border="0"> <tr> <td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td> </tr> <tr> <td>Strongly Agree</td><td></td><td></td><td></td><td></td><td></td><td>Strongly Disagree</td> </tr> </table> <p>5. Comments/suggestions for general appearance</p> <p>_____</p> <p>_____</p> <p>_____</p> <p>_____</p> <p>_____</p> <p>6. Other comments/suggestions for the Virtual Library homepage</p> <p>_____</p> <p>_____</p> <p>_____</p> <p>_____</p> <p>_____</p> <p>_____</p> <p>Thank you for your assistance! Your input will help us in our efforts to continually improve our services to the University community.</p>	1	2	3	4	5	6	7	Strongly Agree						Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree						Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree						Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree						Strongly Disagree
1	2	3	4	5	6	7																																																			
Strongly Agree						Strongly Disagree																																																			
1	2	3	4	5	6	7																																																			
Strongly Agree						Strongly Disagree																																																			
1	2	3	4	5	6	7																																																			
Strongly Agree						Strongly Disagree																																																			
1	2	3	4	5	6	7																																																			
Strongly Agree						Strongly Disagree																																																			

complete. This would show up on the report log and signal the ending of an exercise and the beginning of the next. The concept was tested and found to work adequately. This required an alteration in the instructions for the test to prompt the test participant to navigate back to the homepage after completing each exercise. Once the USG felt that the instrument was ready for testing, it was then time to recruit participants for the first formal test.

Recruitment of test participants

After reading and consulting other papers and case studies on usability testing, a general consensus of recruitment methods began to appear. In most cases, it was stressed that a fairly limited number of participants (usually eight to ten) would serve as a sufficient sample population for testing (Chisman *et al.*, 1999). However, because of the largely unmediated methodology chosen for the USF usability test, it was decided to recruit two groups of 16 participants (32 in all) for participation in the study. The USG recognised early on in the process that unmediated testing methodologies result in a fair amount of unusable data in the form of improperly followed instructions or incomplete tests. The group hoped that the large number of participants would perhaps yield at least ten viable tests out of each group.

Volunteers for participation were recruited from undergraduate classes in the College of Arts and Sciences. In most cases, instructors were contacted and asked to present the idea to their classes and instruct any interested students to contact one of the test administrators. This method was extremely successful and yielded enough participants to complete two rounds of testing. Recruitment fliers were also created and posted, but yielded no volunteers. Many of the participants were new freshman students who had very little, if any, experience using the virtual library. Other participants were upperclassmen who had moderate to extensive experience using the virtual library. Volunteers were released from one hour of class time, but were otherwise uncompensated for participation. It was stressed that their involvement was important to the development of library services and they were thanked for their participation and input.

Testing environment and test administration

The first formal usability test was administered in one of the library's computer laboratories. The laboratory contains 16 individual computers, complete with Internet access. The test participant population consisted of 32 volunteers, divided into two groups of 16 participants. One group participated in the first round of testing during the morning; the other group during the second round in the afternoon. In each case, a test facilitator was present to welcome the participants and provide preliminary instructions for the testing procedure. The facilitator was also present during the testing to answer any questions that arose, or to troubleshoot any problems that might arise with the computer equipment. The facilitator did not directly observe the testers' actions, nor did the facilitator sit close enough to the test participants to give the impression that they were being watched. At the end of the test, the facilitator thanked the test participants and collected their test sheets.

Prior to the arrival of the test participants, an exercise sheet and opinion survey were placed at each workstation. The IP number of the workstation was recorded on each test sheet. When the test participants arrived, they were instructed to select a seat and to wait for further instructions. The computers in the labs come with a "black box" module that allows the facilitator to lock each workstation. This was done to keep the participants from interacting with the interface prior to the start of the test. After all of the volunteers had arrived, the facilitator read through the instructions carefully. The participants were asked if the instructions were clear and if they had any questions. After any questions were answered, the workstations were released and the testing began. Each test participant was asked to copy the time from the lower right-hand corner of the monitor on his/her test sheet before beginning the first exercise. Additionally, each participant was also instructed to record the completion time on the lower part of the test sheet after they had completed the last exercise and before moving on to the opinion portion of the test.

On average, each test participant took approximately 45 minutes to complete both portions of the test. Each participant was excused after completing and handing the test to the test facilitator.

Results

Part I

In the first round of testing, only three tests had to be discarded. One of the participants first clicked on the link for the St. Petersburg regional campus library page and continued to use that page as the homepage for the duration of the test. The reasons for this are still unknown and there is even more confusion as to why the individual continued to return to that page and use it for each exercise. The computer used for that particular test was examined to reaffirm the homepage designation for the browser and investigate any possible tampering with the settings. These all checked out. When the group used the log generated from the participant's test to recreate the tester's actions, it became apparent that the individual used the St. Petersburg regional campus library's homepage to get to the USF virtual library page. The problem was that the virtual library homepage in use at that time was significantly different than the test page being used by the USG. Therefore, this participant's actions did not accurately reflect interaction with the test page and had to be discarded. The other two problem tests involved participants who failed to navigate back to the homepage after each exercise. The exercises on their logs were impossible to differentiate from each other.

In the second group, only one test had to be discarded. The error on this test was quite telling and prompted the USG to think a bit harder before engaging in the next round of testing. In this instance, the test participant did indeed click on "home" after completing each exercise. The problem was that the individual used the "home" link on the virtual library's navigational toolbar rather than the Web browser's "home" link as instructed. The action brought the participant to the virtual library homepage that was currently in use rather than the test homepage. The fault lies not in the participant's actions, but rather the test designer's failure to foresee this possibility.

After the first round of testing was complete, the usage logs were generated and printed out for analysis (see Figure 6) The results were at once revealing and frustrating. It became immediately obvious that the interface was not intuitive, and perhaps even misleading. Indeed, 100 percent of test participants (not counting

discarded results) failed to navigate successfully to the proper location in response to the request to "Locate the link(s) you would click on to research journal or magazine articles". In the first group of participants, 12 out of 12 chose the link labelled "E-journals" rather than the proper "Databases" (see Figure 7). On reflection, however, the participants' actions were the most logical in response to the question. The term "databases" means little to someone unfamiliar with library jargon. E-journals, however, seems quite clear, even though it was not the correct answer. In other cases, the results showed failures in the interface design that appeared to the usability group to be quite obvious. For example, almost every participant missed the correct link to navigate to the main university homepage, even though the university mandates that the university logo and link be placed in the upper left portion of every main page.

Most of the exercises required only one click to complete. However, those that required two or more navigational clicks were revealing in that the participants uniformly navigated to the first location, but failed to go any further to complete the exercise on those that required two or three clicks (see Figure 8).

Other questions showed a high rate of success. Of the participants 92.3 percent correctly located the proper link for renewing library materials online (see Figure 9). Likewise, 76.9 percent successfully located the online help sheet for finding articles (see Figure 10). Overall, out of 15 exercises, the majority of test participants failed to navigate successfully to the requested resource in nine out of 15 exercises, or 60 percent of the time (see Figure 11). Taking the total number of incorrect responses out of the total number of possible correct responses for each test participant, there were 119 incorrect responses and 76 correct responses out of a possible 195 correct responses, showing 1.6 incorrect answers for each correct response.

Part II

For the second portion of the test designed to elicit general impressions of the interface design, more than half of the participants (53 percent) either strongly agreed or agreed that they found the design visually pleasing. Almost half of the participants (40 percent) reported that they found the category headings and links easy to understand and

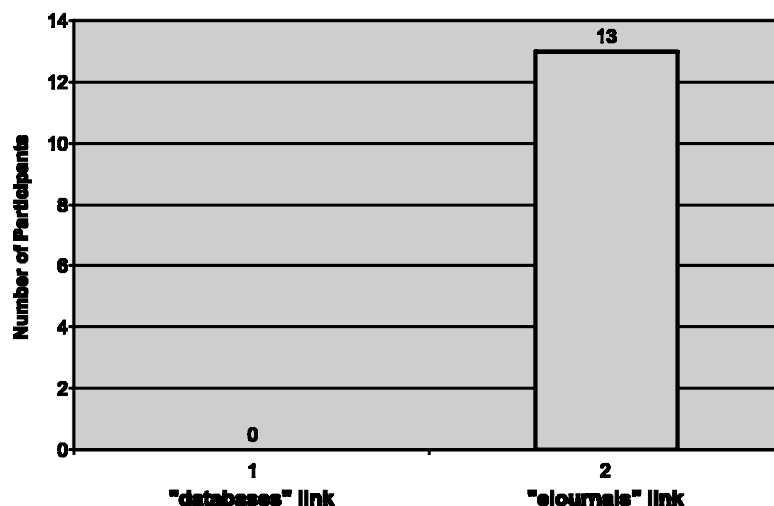
Figure 6 Sample log report

```

1 lab44 - - [21/Mar/2001:09:42:30 -0500] "GET /virtual/newspapers/index.php?db-A HTTP/1.0" 200 19179
2 "https://www.lib.usf.edu/virtual/indextest.html" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
3 lab44 - - [21/Mar/2001:09:43:25 -0500] "GET /virtual/resources/morecats.html HTTP/1.0" 200 6703
4 "https://www.lib.usf.edu/virtual/indextest.html" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
5 lab44 - - [21/Mar/2001:09:43:59 -0500] "GET /proto/samples/articles_db_es.html HTTP/1.0" 200 1821
6 "https://www.lib.usf.edu/virtual/indextest.html" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
7 lab44 - - [21/Mar/2001:09:44:53 -0500] "GET /virtual/services/online renewal.html HTTP/1.0" 200 7174
8 "https://www.lib.usf.edu/virtual/indextest.html" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
9 lab44 - - [21/Mar/2001:09:46:15 -0500] "GET /virtual/indextest.html HTTP/1.0" 200 8345 "-" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
10 lab44 - - [21/Mar/2001:09:46:40 -0500] "GET /virtual/help/book.html HTTP/1.0" 200 7662 "https://www.lib.usf.edu/virtual/indextest.html"
11 "Mozilla/4.08 [en] (WinNT; U ;Nav)"
12 lab44 - - [21/Mar/2001:09:47:11 -0500] "GET /virtual/reftools/ HTTP/1.0" 200 7575 "https://www.lib.usf.edu/virtual/indextest.html"
13 "Mozilla/4.08 [en] (WinNT; U ;Nav)"
14 lab44 - - [21/Mar/2001:09:47:53 -0500] "GET /virtual/indextest.html HTTP/1.0" 200 8345 "-" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
15 lab44 - - [21/Mar/2001:09:48:03 -0500] "GET /virtual/help/article.html HTTP/1.0" 200 6212
16 "https://www.lib.usf.edu/virtual/indextest.html" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
17 lab44 - - [21/Mar/2001:09:48:41 -0500] "GET /virtual/indextest.html HTTP/1.0" 200 8345 "-" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
18 lab44 - - [21/Mar/2001:09:49:21 -0500] "GET /virtual/services/all.html HTTP/1.0" 200 5879
19 "https://www.lib.usf.edu/virtual/indextest.html" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
20 lab44 - - [21/Mar/2001:09:49:27 -0500] "GET /virtual/indextest.html HTTP/1.0" 200 8345 "-" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
21 lab44 - - [21/Mar/2001:09:49:55 -0500] "GET /virtual/resources/usflinks.html HTTP/1.0" 200 6250
22 "https://www.lib.usf.edu/virtual/indextest.html" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
23 lab44 - - [21/Mar/2001:09:50:24 -0500] "GET /virtual/indextest.html HTTP/1.0" 200 8345 "-" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
24 lab44 - - [21/Mar/2001:09:50:28 -0500] "GET /virtual/help/gettingstarted.html HTTP/1.0" 200 11167
25 "https://www.lib.usf.edu/virtual/indextest.html" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
26 lab44 - - [21/Mar/2001:09:50:36 -0500] "GET /virtual/personalize/ HTTP/1.0" 200 5320 "https://www.lib.usf.edu/virtual/indextest.html"
27 "Mozilla/4.08 [en] (WinNT; U ;Nav)"
28 lab44 - - [21/Mar/2001:09:50:50 -0500] "GET /virtual/help/gettingstarted.html HTTP/1.0" 200 11167
29 "https://www.lib.usf.edu/virtual/indextest.html" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
30 lab44 - - [21/Mar/2001:09:51:01 -0500] "GET /virtual/libraries/ HTTP/1.0" 200 5843 "https://www.lib.usf.edu/virtual/indextest.html"
31 "Mozilla/4.08 [en] (WinNT; U ;Nav)"
32 lab44 - - [21/Mar/2001:09:51:11 -0500] "GET /virtual/resources/usflinks.html HTTP/1.0" 200 6250
33 "https://www.lib.usf.edu/virtual/indextest.html" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
34 lab44 - - [21/Mar/2001:09:52:00 -0500] "GET /virtual/indextest.html HTTP/1.0" 200 8345 "-" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
35 lab44 - - [21/Mar/2001:09:52:19 -0500] "GET /virtual/resources/usflinks.html HTTP/1.0" 200 6250
36 "https://www.lib.usf.edu/virtual/indextest.html" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
37 lab44 - - [21/Mar/2001:09:55:57 -0500] "GET /virtual/indextest.html HTTP/1.0" 200 8345 "-" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
38 lab44 - - [21/Mar/2001:09:56:13 -0500] "GET /virtual/browse/ HTTP/1.0" 200 11256 "https://www.lib.usf.edu/virtual/indextest.html"
39 "Mozilla/4.08 [en] (WinNT; U ;Nav)"
40 lab44 - - [21/Mar/2001:09:56:40 -0500] "GET /virtual/indextest.html HTTP/1.0" 200 8345 "-" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
41 lab44 - - [21/Mar/2001:09:57:01 -0500] "GET /virtual/help/proxy.html HTTP/1.0" 200 7741 "https://www.lib.usf.edu/virtual/indextest.html"
42 "Mozilla/4.08 [en] (WinNT; U ;Nav)"
43 lab44 - - [21/Mar/2001:09:57:37 -0500] "GET /virtual/indextest.html HTTP/1.0" 200 8345 "-" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
44 lab44 - - [21/Mar/2001:10:05:47 -0500] "GET /virtual/indextest.html HTTP/1.0" 200 8345 "-" "Mozilla/4.08 [en] (WinNT; U ;Nav)"
45 lab44 - - [21/Mar/2001:10:05:48 -0500] "GET /virtual/ HTTP/1.0" 200 8368 "-" "Mozilla/4.08 [en] (WinNT; U ;Nav)"

```

Figure 7 User response to question 2: "Locate the link(s) you would click on to research journal or magazine titles"



navigate. The same percentage claimed that they found the homepage easy to read. Given the poor performance of the interface in the navigational portion of the test, these results were somewhat confusing.

Making changes – testing again

From the information gleaned from the first round of testing, the following changes were made in the interface to conform more to user behaviour: The two individual links for “E-journals” and “Databases” were combined into one link labelled “find an article”. The new link referred the user to a new intermediate page where the option for

choosing direct subscriptions to e-journals and those articles found in databases were offered and explained. In light of the test participant behaviour in the previous round of testing, whereby the full path to a resource that required more than one navigational click went unrealised, the group understood that the collapse of two linked resources into one with an intervening page might result in an equally unsatisfying result. However, upon examination of the language used on the original test page, the simple, straightforward approach as seen in the “how do I” questions invariably produced more successful participant behaviours than those using

Figure 8 Required number of clicks versus number of recorded clicks

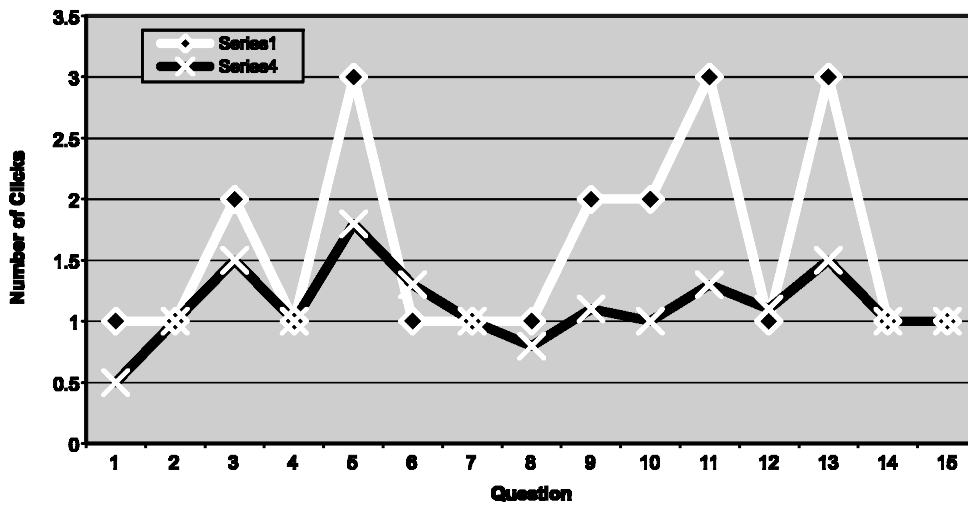


Figure 9 User response to question 2: "Locate the proper link(s) for measuring library materials online"

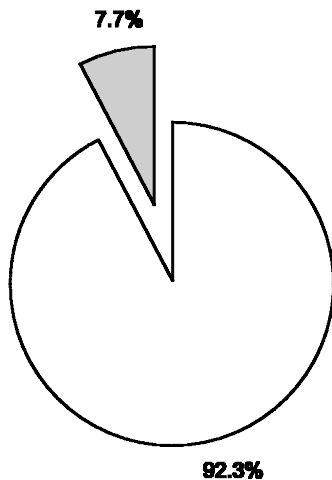
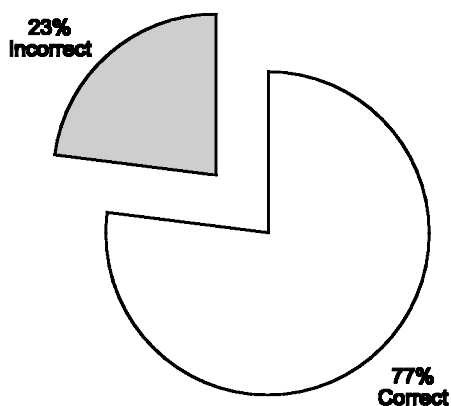
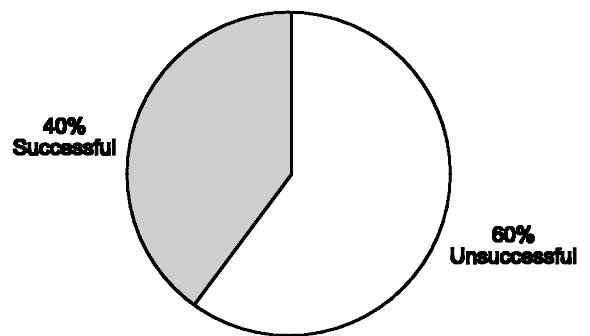


Figure 10 Percentage response to question 6: "Locate an online help sheet that will tell you how to find an article"



jargon. The group gambled on the idea that the link labelled "find an article" would be more successful, even if the test participant were confronted with two additional choices

Figure 11 Overall percentage of successful vs unsuccessful navigations in round 1



after selecting that link. In the same vein, the link for the online catalogue was changed from "WebLUIS" to "find a book". The explanatory text "Go to the USF Homepage" was added to the graphical USF logo/link in the upper left portion of the page and the font size of the links to the other library locations was made larger (see Figure 12).

In addition, some changes were made in the wording of the test instructions and questions to make them more explicit in the hope that fewer participants would skip questions, or complete exercises out of order.

Once these changes were completed, the second group of participants was recruited and the testing conducted again.

Second round results

Test participants in the second round of testing fared quite a bit better than the first group.

More than half of the participants failed to navigate to the correct locations only 43 percent of the time (see Figure 13). Out of a

Figure 12 Explanatory text and larger fonts

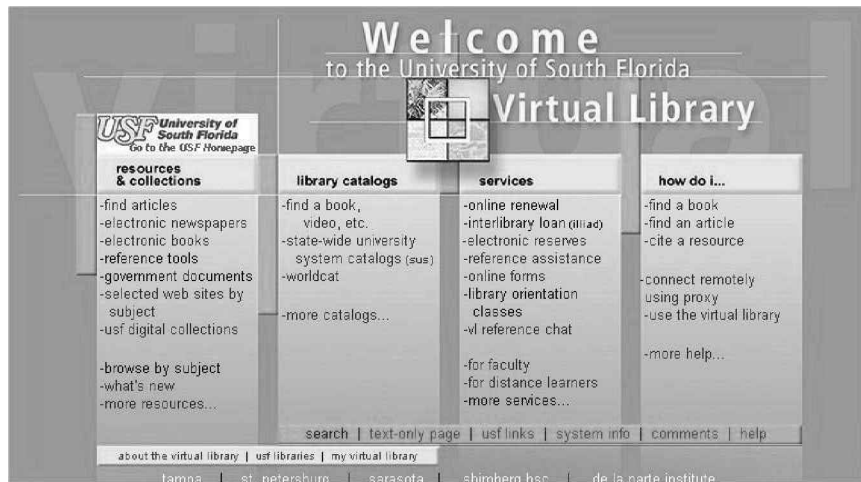
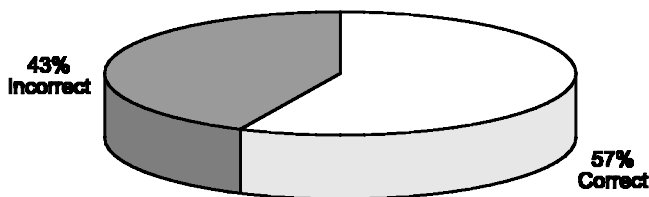


Figure 13 Overall correct vs incorrect responses on round 2 of testing



possible 225 correct answers, participants navigated correctly 128 times for an overall success ratio of 7.6 correct answers for every incorrect answer. This translates into 57 percent correct answers. This represents a marked improvement over the results of the first test. Interestingly, the errors made by the participants on the second test were quite different than those of the first. For instance, rather than having trouble locating the proper resources for finding journal articles or materials in the library, most participants demonstrated a significant amount of difficulty in locating the requested “search” feature to launch a search of the World Wide Web (see Figure 14). The majority of participants clicked on the link labelled “how do I . . . use the virtual library”, most likely in an effort to locate further assistance. The actual “search” link, located on a narrow bar toward the bottom of the page, went largely unnoticed and unused by most of the participants in the second round of testing. Curiously, this was not such an issue with the first group, even though this link did not experience a change in location, colour, size or language.

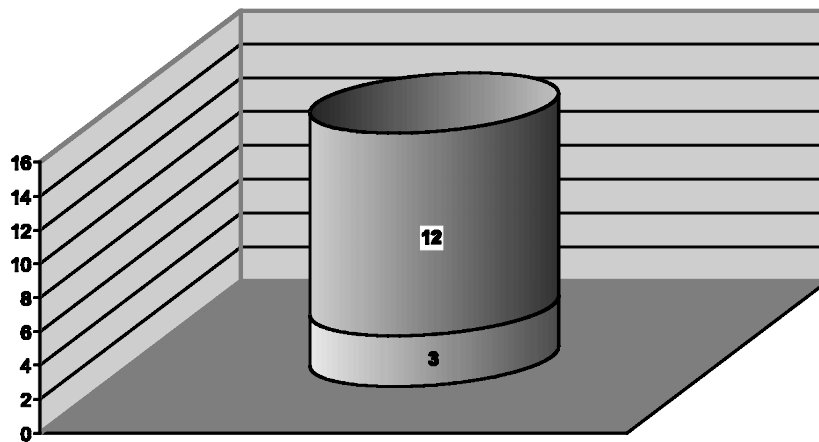
For the second portion of the test, 65 percent strongly agreed or agreed that the

interface was visually pleasing. A total of 42 percent reported that they found the categories and links easy to understand and navigate; 65 percent agreed that the category headings were easy to understand.

Most interesting was the apparent discrepancy in both testing sessions between test participants’ actual performance and their perceived performance, reflected in the feedback on the second part of the test. Most participants rated the interface quite favourably for its navigability and visual aesthetics, yet the results show a large number of missed answers; needlessly long navigational paths, and fairly long intervals of time between mouse clicks, clearly indicating user difficulty. This would suggest that either the test participants are reluctant to give negative feedback, or they simply do not understand the virtual library concept well enough to know when they are experiencing difficulties. If it is the former, then test questions of this type must be considered relatively useless and should not be included in usability testing. If the latter is indeed the case, then more aggressive information literacy programmes are needed.

Conclusions and observations

One strong indicator that more alterations and testing are needed lie in the number of navigational clicks to individual resources recorded for each user. In most cases, this number was consistently above the necessary number of clicks required to navigate successfully to each requested

Figure 14 Question 11: "Locate the link(s) that will take you to a place where you can search the World Wide Web"

resource. This remained consistent from the first round of testing to the second. Second, it became increasingly obvious that plain, straightforward language almost always produces better results than using jargon. Unfortunately, resistance among those responsible for changing this language has been higher than expected. Even though testing has clearly demonstrated problems, the response to correct the interface has been sluggish. Many of the recommended changes have gone unimplemented. The present interface that is currently being used reflects the new design and is easier to navigate than the older one, but still contains quite a few identified problem areas, the majority dealing with the language used for link labels.

Informal consultation with members of the library faculty has provided some insight as to why these changes have gone unrealised. Overwhelmingly, librarians cite an extreme reluctance to change an interface with which they have become familiar. Interestingly, many others even expressed surprise when informed that the usability testing exposed problems with the design; it simply had not occurred to them that users would not understand much of the terminology currently in use. One particularly singular comment argued that library jargon must not be eliminated from user interfaces, but that the library must make a more concerted effort to educate its user base regarding the meaning of these terms. In some ways, this educational approach that emphasises the introduction of the user to the lexicon of librarianship is admirable. But we must question the practicality and feasibility of this kind of mission.

In the world of academia, we are subject to an influx of new users each semester. Invariably, these new users arrive unprepared to delve into the world of scholarly online information retrieval. For the most part, these individuals are familiar with a small number of Web search engines and directories. They are not familiar with terms such as "online catalogue", "databases", or "e-journals". Moreover, a substantial portion of student users rarely visit the library at all, preferring to access provided resources remotely. The library's education programme, while extremely active and growing, still fails to reach every patron. In tandem with these concerns are those that incorporate the community outside the university and the library's obligation to society as a publicly funded institution. Considering all the components that factor into the decision to either retain and educate, or eliminate jargon from an online user interface, it would appear that the former is a daunting mission that would meet with little success. One must also ask why it is so undesirable to make these concessions when they are in the best interests of usability.

What is next

If any valuable lessons were gleaned from this initial testing effort, the most apparent was that usability testing should be conducted on a continuous basis. What's more, other methodologies overlooked by the group for this study would undoubtedly prove beneficial. The most glaring omission was that of card-sorting techniques. Testing for the new VL interface began with a pre-sorted

and preconceived idea of link categorisation. It would be most interesting to return to a more chaotic mind-set and see how users would organise resources. Research into the field of card-sorting methodologies has revealed some interesting and useful software applications that automatically randomise resource labels and interactively allow the user to group the resources into either pre-named categories, or categories to be named later by the test participant. The two applications reviewed by the test administrator for possible use in future usability tests include the freely available Web Category Analysis Tool (WebCAT) from the National Institute of Standards and Technology (<http://zing.ncsl.nist.gov/WebTools/WebCAT/overview.html>) and the EZSort beta release from IBM (http://www-3.ibm.com/ibm/easy/eou_ext.nsf/Publish/410). Each application requires a significant investment of time for the initial set-up and implementation, but the results obtained from electronic card-sorting techniques are more reliable and less apt to be influenced by the test creators, or subject to physical accidents (such as dropping the sorted card pile).

All in all, the experience of doing usability testing was rewarding as well as revealing. Some results were anticipated while others were a complete surprise. While the intense

fervour over much of the Web and interface design has subsided in recent months, those who provide information services and act as a gateway to hundreds, if not thousands, of resources must be ever diligent to ensure that they are meeting their users' needs and responding to user input. Usability testing represents one important means of accomplishing this goal.

References

- Chisman, J.K., Diller, K.R. and Walbridge, S. (1999), "Usability testing: a case study", *College & Research Libraries*, Vol. 60 No. 6, pp. 552-69.
- Dickstein, R. and Mills, V. (2000), "Usability testing at the University of Arizona Library: how to let the users in on the design", *Information Technology and Libraries*, Vol. 19 No. 3, pp. 144-51.
- Dumas, J.S. and Redish, J.C. (1993), *A Practical Guide to Usability Testing*, Ablex, Norwood, NJ.
- Fichter, D. (2000), "Head start: usability testing up front", *Online*, Vol. 24 No. 1, pp. 79-81.
- Fraser, B., Comden, D. and Bergstahler, S. (1998), "Including users with disabilities: designing library Web sites for accessibility", *CHOICE: Current Reviews for Academic Libraries*, Vol. 35, supplement, pp. 35-7.
- Head, A.J. (1999), "Web redemption and the promise of usability", *Online*, Vol. 23 No. 6, pp. 20-3.
- Rubin, J. (1994), *Handbook of Usability Testing: How to Plan, Design and Conduct Effective Tests*, John Wiley & Sons, New York, NY.