Yeah, I Found It!: Performing Web Site Usability Testing to Ensure that Off-Campus Students Can Find the Information They Need.

Beth Thomsett-Scott
University of North Texas

Abstract

A well-designed and user-friendly web site is essential with the large and rapid increase in the number of off-campus users. Web site usability studies can help ensure that libraries provide effective and efficient access to their web site for both on and off campus users. Formal usability studies, focus groups and cognitive walkthroughs are usability techniques that can be incorporated into a web site study. These three techniques will be discussed with real life experiences added in to provide the reader with suggestions of how to and how not to use the techniques.

Introduction

The number of off-campus users of library sites is rapidly increasing as computers become more prevalent among students and online courses grow in number and popularity. Library web sites are not always designed to allow their patrons to use the site efficiently and effectively without assistance or frustration. Dickstein and Mills (2000) state that a web site must be designed for the users of the site, not for the librarians or the creators. Libraries are beginning to adopt the concepts of "usability" that have been used by software companies for a number of years. The interest in providing web sites that are seen as "useable" and "helpful" by our patrons comes not only from a desire to provide the best possible service but also from other factors. One of these factors is that the increase in distance learning and off-campus users necessitates additional concern about the efficiency and effectiveness of a library's web site (Cockrell & Jayne, 2002). As well, libraries are spending more and more of their budgets on web-based resources and it is essential that users be able to access and use these resources efficiently and effectively.

A useable web site is defined by Rubin (1994) as being useful (doing what the user wants it do), effective (the ease of use while achieving the user's goals), learnable (the ability to grow from a novice to a skilled user) and satisfying (how well the user enjoyed their time on the site). McMullen (2001, p.7) states that a useable web site is essential as it is the interface between the "information professional and the individual who is seeking information". Usability studies gather information on how users actually use the site (Campbell, Walbridge, Chisman, & Diller, 1999). Battleson, Booth, and Weintrop (2001) state that usability studies replace "opinion" with user-centered data. Usability studies put the end users first and assist in understanding the needs of the users (Norlin & Winters, 2002). Collins and Aguinaga (2001) describe their amazement at how differently the study participants used the site than the staff or the designers of the site expected it to be used. Incorporation of usability testing into web site development increases the value of the site and enhances user satisfaction (Norlin & Winters, 2002).

The purpose of this paper is to discuss several web site usability techniques using real life experiences to provide the fundamentals of the techniques and to help prepare the reader to use them in a web site usability study. The author's experience is with academic libraries; however, web site usability testing can be used on any web site. Some knowledge of web site usability is assumed. For readers who are unfamiliar with the topic, Campbell (2001a) and Norlin and Winters (2002) are both good sources for background information.

Background to the real-life studies

Two web site usability studies were conducted at the University of North Texas (UNT) Libraries. The first study was performed on the Ask A Librarian sub site and the second on the home page. Thirty

students were hired for each study. The participants were divided into three categories: undergraduates, graduates and those experienced with online learning.

Both studies used the techniques of formal usability studies and focus groups. The formal usability study technique was incorporated as it involves users directly in the process and was performed first as the results formed the basis of the discussion in the focus groups. Jeffries and Desurvire (1992) suggest that the best evaluations come from multiple methods. Focus groups were included as they provide group discussion and often promote the synergy of ideas. Significant additional information was gleaned from the groups in both studies. Often participants are more relaxed in a group situation than in one-on-one situations (Canning, 1995), and this may be the reason for the additional usefulness of the focus groups.

Formal Usability Studies

The Basics

Formal usability studies, also referred to as task-based testing, are the ultimate way to involve library users in the design of the site. They allow researchers to observe how the site is used, find any errors in the design, and determine which elements or terminology participants find confusing. Since formal usability studies provide both qualitative and quantitative data, they give greater scope and breadth to the study (Norlin & Winters, 2002). Formal usability studies find the more serious difficulties that users have with the site (Jeffries & Desurvire, 1992).

Formal usability studies involve the observation of users performing specific tasks on a web site. Observation can be direct where a moderator, and sometimes an observer as well, watches the participants while in the same room; or indirect using a two-way mirror, video recorder and/or screen capture equipment. It is usually only the bigger companies or institutions that have two-way mirrors and the more advanced set-ups. Most research is still done using direct observation with the addition of screen capture software. Screen capture software, often called tracking software, collects information on the users' paths while they complete the tasks. Basic screen capture software will record the time taken between each mouse click and track the URLs of the pages viewed. More advanced systems, and more costly, will record screen images and some even produce a video of the transaction.

There are usually eight to twelve questions with very specific answers and paths, as the goal of usability studies is to determine how the individual pages and overall site can be improved to allow users to find the information they need in less time and with fewer steps. The tasks are based on the actual purposes of the site, such as finding the catalog, locating the hours page, finding a liaison librarian etc. The difficulty in writing the tasks is avoiding using the terminology that is actually on the page. Tasks are the most critical part of the design (Campbell, 2001b), and need to be solvable without "leading" the participant through the use of keywords from the "correct" link. The wording and the order of the tasks are the keys to a successful study (Battleson et al., 2001). Chisman, Diller & Walbridge (1999) suggested that tasks should not be built on one another as this will prevent unsuccessful students from answering the next question as well.

Each test session usually lasts between one and a half hours and two hours. The data collected can vary among researchers; however, most record the success rate, the time taken to complete each task, the number of clicks, and the pathway. The pathway is useful in determining the common mistakes that participants make. Some researchers will record body language observations as well to help them determine at what point the participants become frustrated with the question.

There is some debate about timing the sessions, both whether it should be done and how long to do it for. Some researchers do not time at all and allow the participants to continue searching for the answer as long as they want to. Other researchers stop the participants at one minute as they believe that if the participant is not successful after one minute, they will not be able to find the correct answer at all. However, the standard still appears to be a three minute maximum.

Task-based testing can be as minimalistic or as extensive as needed. Rubin (1994) states that four to five participants will find eighty percent of the errors and fifteen will find all the errors at the page level. Participants must be homogeneous for the above to be true. If differences among populations are being considered, more participants will be needed to ensure a minimum of four to five homogenous participants in each population. Skelton (1992) notes that any changes based on a single test can actually worsen the site. Therefore, after applying the results of the usability study to a redesign, iterative testing, which involves the testing the changes in a less formal atmosphere before revealing them to the public, should be performed to ensure that the study results were interpreted correctly (Campbell et al., 1999). This process is also called “cookie testing”.

Comments and Thoughts

Dickstein and Mills (2000) used eight to twelve participants across several populations and noted that the first four to five participants found the major errors, but added that additional participants always revealed something valuable and unique. Each of the ten participants in each population in the UNT studies provided some useful information, although the more “serious” errors were discovered after the fourth to sixth participant in each population.

There were eight questions in the first study and ten in the second. Ten questions are recommended as the maximum number to use. After the tenth question, participants were showing signs of fatigue and becoming more easily frustrated.

Beta testing the tasks prior to the study is highly recommended. If different populations are being examined, the beta testing needs to be performed across all the populations. Hudson (2002) discussed the importance of beta testing across populations as well. An entire beta test should be conducted to determine any potential pitfalls so they can be resolved prior to the main study. Data collection and analysis should also be beta tested to ensure that the necessary data will gathered properly and entered in a manner to facilitate successful analysis.

Both a moderator and observer were used when possible. The moderator’s role was to introduce the test, ask the participants to read the questions aloud, and prompt them if they left the site. The observer would time the test and record the pathway taken for each question. Approximately a third of the second series of tests only had a moderator present due to unavailability of volunteers to serve as recorders; thus, the moderator was required to also act as the observer.

The time allotted for each participant to complete all the tasks was one and a half hours, although both series of tests ended before that time. However, the extra time allowed a relaxed debriefing period and casual conversation, if the participants desired this, which added to the overall positive feeling that a number of the participants commented on.

Three minutes was the maximum time allowed for each task and this worked well. Only two participants became frustrated enough that the question had to end prior to the three minutes. In both cases it was the same question that caused the problem. It has been discussed in the literature that if the participants cannot find the correct path after one minute, they would not be successful at completing the task (Gullikson, Blades, Bragdon, McKibbin, Sparling, & Toms, 1999); however, the results do not support this. In fact, there were a number of times when the question was successfully answered just before the three minute mark, although three minutes did seem to be the longest participants wanted to, or were able to, look for the answer. Battleson et al. (2001) allowed three minutes as the outside marker for closing the question, but used three incorrect choices as the time to intervene.

Participants were asked to read the questions aloud, as suggested by Battleson et al. (2001), to encourage the participants to continue to think aloud while completing the question. However, most participants did not vocalize unless they were reading the question. Walbridge (2000) notes that the think-aloud technique reveals the thought processes of the participants, which can indicate why they make particular choices. It is recommended that usability participants be encouraged to think aloud as the information obtained should be well worth the effort.

A video capture software was used which provided high quality videos, and allowed the researchers to go back and check any pathway or timing questions they had. McGillis and Toms (2001) used tracking software exclusively. The video capture software was particularly important in the UNT studies due to the high number of sessions where only a moderator was present.

Participants were allowed to choose either Netscape or Internet Explorer to complete the tasks. McGillis and Toms (2001) suggest that participants be allowed to use the browser of their choice to avoid any variation in usability caused by an unfamiliar browser. Many students were content with Netscape, the default on the terminals, but others appreciated the ability to use the browser they were most comfortable with.

Tasks were printed out and available for the participants to look at while doing the study. Participants appreciated having the questions in front of them, although it had to be made clear several times that participants were not expected to write the path on the sheets! A number of students would pick the question up and look at it or glance down at it. Chisman et al. (1999) notes that non-English speakers had difficulty understanding some of the test questions, thus supporting our decision to provide the questions in written form as it is frequently easier to understand information presented in written form than in verbal form.

At the end of the session, participants were offered a chocolate bar or box of candy to provide “immediate gratification”. Participants appreciated the chocolate bars as some were “down” after the session and the candy perked them back up. This is highly recommended as the immediate gratification effect seemed to overcome any negative feeling from the study. It is important to have participants feel good about the process, especially if you want them to return for another part of the study or to be available for the next study.

A post-test questionnaire, which asked the participants for their likes and dislikes of the site as well as what made it easy and difficult to use, was administered after the formal usability test. The post-test questionnaire elicited a lot of useful information. Most students did not fill in all the blanks and this was permitted as forcing them to fill in all the blanks would make it more like a “test”. It was more important that the participants left with a good feeling about the library. Often the comments reflected what had been observed, so overall there wasn’t as much data missing as the empty blanks would suggest. McGillis and Tom (2001) note that post-test surveys provide the participants’ perceptions of using the site and the features. Questionnaires are useful to balance the feedback from the study (Campbell, 2001a). Participants were also asked to fill in a demographic questionnaire.

The use of a post-test questionnaire allowed students some time to relax and led into the debriefing sessions. Debriefing sessions for the moderator and observers were also included as time permitted. Chisman et al. (1999) state that debriefing participants and test administrators after each study is essential for a quality result. Walbridge (2000) notes that debriefing offers an opportunity to provide information literacy training to the participants. Debriefing also allows the moderator and observers to examine the results. Rubin (1994) notes that the debriefing portion is the key to fixing the problems indicated by the study results. The informal comments provided by the participants during the debriefing were often enlightening (Battleson et al., 2001), possibly because the “test” had concluded.

Data Recording, Analysis and Presentation of Results

A results sheet was used for each participant and included the names of the moderator and observer, the time taken to complete the task, whether the task was successfully completed, what pathway was taken to complete the task, and what problems were encountered. The observers were encouraged to note the behavior of the participants and to include any comments that were expressed as part of the “think-aloud” process on the sheet. Recording the pathway, as well as the number of mouse clicks or pages viewed, provides information on common mistakes which helps to determine what items need to be changed to make the pages more useable. Participants were required to return to the home page before beginning the next question in order to provide a standard starting point for each question.

Using Excel, the success rate, time taken, and number of clicks were recorded for each question across all participants. The average, mode and median were calculated. Pathway information was recorded in a Word table using the name of the link the participant chose. Post-test questionnaire responses were also recorded in a Word table. The participant number was attached to all data.

The Excel data provided quantitative information on how difficult the web site was to use and which tasks were more difficult. The pathways were examined to see where the students took a “wrong turn”. This information was correlated to determine which links or terminology needed revising. The post-test questionnaire provided additional insight into the usability of the web site.

Results were presented with suggestions for improvement based on common errors and page problems found during the study.

Focus Groups

The Basics

The focus group technique has been used for many years in marketing, psychology, and sociology research. Focus groups are an inquiry method that looks at users’ opinions or perceptions. Libraries use focus groups primarily for user satisfaction studies, although they are beginning to use them to determine marketing strategies and to incorporate them into web site redesign projects. Canning (1995) described focus groups as a cost-effective and quick means to elicit relevant information about services and resources. Focus groups can also be part of web site usability studies.

When incorporated into usability studies, focus groups are used in several different ways. Groups can be used to look at a site initially, usually copies rather than live sites, and discuss issues about the site. This preliminary information is used to create prototypes for the revised site. Other researchers use focus groups after formal usability studies. Some researchers will perform the formal usability studies in a group setting and then immediately have a focus group (Palmquist, 2001), while others work with the usability participants individually and then bring them together for a focus group.

When used in usability studies, focus groups are conducted similarly to other disciplines. There are usually eight to twelve participants, a moderator and perhaps an observer, it may be audio and/or video taped, have eight to ten questions, and last for approximately two hours. Focus groups are another relatively cheap way of gathering input from users as they can be small, with one or two groups, or large, with several to many groups.

Thoughts and Comments

The UNT focus groups ranged from four to six participants. The small size of the groups worked well, especially as participants were allowed to digress from the questions. The researchers believed that this promoted a more positive feeling for the experience and also, for the same reason, would answer very quick, specific reference questions. An example of this process was several participants asked why UNT did not have an online form for interlibrary loans. The moderator replied that there was a form, quickly discussed where the form was located, and used the question as an example of why the studies were being performed! Responding to some questions at the point of need increased participant satisfaction and showed that the Libraries were interested in students needs and not just in gathering data.

Participants were asked for their first comment in a round robin manner to create a comfort level among the group members. After the first set of comments, participants were allowed to offer comments in free form. Generally, after each person had given an answer, there was more interaction among the attendees. Students were respectful of others’ opinions as had been requested at the beginning of the group.

When using focus groups after formal usability studies, an internet accessible computer available should be available or screen shots of the main pages should be provided. Frequently, participants wanted

to review the pages to remember specific comments they had had during the task completion. Screen shots were included in the second series of focus groups and the participants provided many more useful and specific comments.

Another very important lesson learned from the UNT studies was to ensure that the focus groups were held in a welcoming environment. The first set of focus groups was held in a lecture-like room with dark brick walls and it was a frequently used room so the researchers were unable to “decorate” it with homey touches. Participants who had been outgoing and verbal in the usability studies were quieter during the focus groups, which is the inverse to the expected situation. The second series of focus groups was held in the same room as the usability studies which, although significantly smaller, worked very well. Even participants who were subdued during the usability studies were comfortable during the focus groups. Collins and Aguinaga (2001) note that they took great care to make the test room homey and comfortable.

Participants were provided with comment sheets that had each question written on them, with one question per page. The sheets were introduced by the moderator as a place to write down comments participants didn’t have time to make during the discussion or comments they didn’t want to vocalize. Pens were provided to avoid any concerns that their comments could be traced back to them as anonymity had been assured. The comment sheets were generally not utilized, although there were several occasions where students used them when they did not agree with the general consensus of the group and wanted to ensure that their comments were recorded. Since the groups were small and there was plenty of time allowed, it is possible that students didn’t have many more comments to add. Comment sheets will be provided in subsequent studies as the few comments received were very valuable.

Data Collection, Analysis and Presentation of Results

Data collected from focus groups can be difficult to analyze as it is primarily qualitative information rather than quantitative. One method to deal with this is to put the comments in a Word table. Then, assign a subject heading or category to each comment. Using this method makes it easy to gather comments with the same subject or category together using the sort command in Word. Another way is to enter all comments, scan them for trends, decide on major categories or subjects, and cut and paste comments into the appropriate category. The number of positive and negative comments in each category can then be counted to gather some quantitative information on each category. Comments are broken into very small pieces using this technique. Participant numbers or identification should be assigned to each comment to track the owner of the comments especially if trends across different populations are being looked for.

Examples of categories include design issues, such as colors and navigation; comments on specific pages; terminology; and desired links - links users would like to see on the home page or other pages. Categories will vary depending on the questions asked and the discussion threads. A study by Thomsett-Scott et al (unpublished), utilized both focus groups and formal usability studies to explore the best and the worst of the UNT Libraries’ home page and three other academic library home pages. During the focus groups, the participants pulled out links and terminology that were missing for the UNT page and discussed design elements, such as a direct catalog search box on the home page, which would improve the usability of the UNT home page.

Cognitive Walkthroughs

The Basics

The cognitive walkthrough is an inspection technique where “usability experts” or designers work through the web site using scenarios that approximate the users’ experiences. The scenarios are often very complex and require the walkthrough team to completely get into the users’ point of view. For example, a common scenario is “I am a first year student and need to find three magazine articles on anemia. I have never used the library before and don’t know what anemia means”. In a face to face situation, the librarian would initiate a reference interview and guide the student through the steps necessary to find information

on their topic. Unfortunately, this student is not in the library but on the internet approaching an unfamiliar and possibly complex web site.

Although the word “expert” is included in the previous paragraph, a well-prepared and user-passionate team of library staff can do almost as well. The essential characteristic of an effective cognitive walkthrough team is being able to look at a question and find the correct route to locate the answer from a user’s perspective (Campbell, 2001a). There are usually two to three members on the team.

There are often eight to ten tasks, frequently based on the most common or essential tasks that the web site is able to answer (Campbell, 2001a). These tasks often necessitate finding the catalog, databases, interlibrary loan forms, contact information, online reference services, subject guides, tutorials and more. The age and library experience of the user are identified as well as any special circumstances, such as a disability.

The team will go through the steps involved to complete each task. As they move through the web pages necessary to answer the task, a list of the pages and the errors or usability issues discovered will be kept. Issues can be direct errors, such as wrong or outdated information, or indirect, such as terminology issues (library speak, multiple terms for one item), and design issues (links not changing color, poor background colors etc.). Essentially, the team needs to look for anything that inhibits the user’s ability to locate the information they need. Required changes can then be prioritized. Some of these will be easy to fix, while others will require additional effort, such as needing the agreement of colleagues when deciding whether the catalog should be titled “online catalog”, “library catalog”, “catalog of books and more” etc. Other “big deal” modifications include changing titles of pages which may affect navigation elements, color standardization among different departments of the library, re-arranging the information or links on pages, and similar issues.

Thoughts and Comments

Cognitive walkthroughs were not selected for the UNT studies as direct user participation was desired. However, they are highly recommended, especially before the formal usability study technique. The benefit of performing a cognitive walkthrough is that the usability team will catch many of the more obvious errors that detract from the usability of the site. While observing participants during the formal usability studies, the UNT researchers discovered two incidences of this type of error: the response time to e-mail questions was listed as twenty-four, forty-eight and twenty-four to forty-eight hours on various pages and several of the links on the home page of the sub site did not change color after being clicked while the remaining ones did. These two errors occasionally distracted participants during the task-based tests. It is worthy of note that the site was examined six months prior to the study. This suggests that reviewing the site without considering the user will find some issues, such as outdated information, grammar and spellings mistakes, but will not catch all of the usability mistakes.

Cognitive walkthroughs are a relatively cheap method of cleaning up a web site. Since walkthroughs don’t involve actual users, they can be done at any time and as frequently as desired. A simple way to get volunteers is to solicit several librarian friends, order in pizza, develop some scenarios and go for it! As mentioned above, the primary qualification for the “walkers” is the ability to put themselves in the user’s shoes. However, there is no substitute for involving the users as even the best cognitive walkthrough team will still miss things that usability studies with real users will find (Campbell, 2001a) .

Summary

With the rapid and still-growing increase in the number of users accessing library resources from off-campus, it is more important than ever that libraries have web sites that are “useable” without library staff mediation. Web site usability testing will help to achieve this. Task-based testing, focus groups and cognitive walkthroughs are relatively easy and effective web site usability testing methods that can be modified to fit most library budgets. Learning from the results of previous studies can increase the

usefulness of the next study. Libraries serving off-campus users will find user satisfaction with their web site increase through the use of web site usability testing.

References

- Battleson, B., Booth, A., & Weintrop, J. (2001). Usability testing of an academic library web site: a case study. *Journal of Academic Librarianship*, 27(3), 188-198.
- Campbell, N. (2001a). Usability methods. In N. Campbell (ed.), *Usability assessment of library-related web sites*, (pp. 1-10). Library and Information Technology Association : Chicago.
- Campbell, N. (2001b). Conducting a usability study. In N. Campbell (ed.), *Usability assessment of library-related web sites*, (pp. 11-15). Library and Information Technology Association : Chicago.
- Campbell, N., Walbridge, S., Chisman, J., & Diller, K.R. (1999). Discovering the user: a practical glance at usability testing. *Electronic Library*, 17(5), 307-311.
- Canning, C.S. (1995). Using focus groups to evaluate library services in a problem-based learning curriculum. *Medical Reference Quarterly*, 14(3), 75-81.
- Chisman, J., Diller, K., & Walbridge, S. (1999). Usability testing: a case study. *College & Research Libraries*, 60:6, 552-69.
- Collins, K. & Aguinaga, J. (2001). Learning as we go: Arizona State University West Library's usability experience. In N. Campbell (ed.), *Usability assessment of library-related web sites*, (pp. 16-29). Library and Information Technology Association : Chicago.
- Dickstein, R. & Mills, V. (2000). Usability testing at the University of Arizona Library: how to let users in on the design. *Information Technology and Libraries*, 19(3), 144-151.
- Gullikson, S., Blades, R., Bragdon, M., McKibbin, S., Sparling, M., & Toms, E.G. (1999). The impact of information architecture on academic web site usability. *Electronic Library*, 17(5), 293-304.
- Hudson, L. (2000). Radical usability (or, why you need to stop redesigning your website). *Library Computing*, 19(1/2), 86-92.
- Jeffries, R. & Desurvire, H. (1992). Usability testing versus heuristic evaluation: was there a contest? *SIGCHI Bulletin*, 24(4), 39-41.
- McGillis, L. and Toms, E.G. (2001). Usability of the academic library web site: implications for design. *College and Research Libraries*, 62(4), 355-368.
- McMullen, S. (2001). Usability testing in a library web site redesign project. *Reference Services Review*, 29(1), 7-22.
- Norlin, E. and Winters, C. (2002). *Usability testing for library web sites*. American Library Association : Chicago.
- Palmquist, R.A. (2001). An overview of usability for the study of users' web-based information retrieval behavior. *Journal of Education for Library and Information Science*, 42(2), 123-136.
- Rubin, J. (1994). *The handbook of usability testing: how to plan, design and conduct effective tests*. New York : Wiley.
- Skelton, T. (1992). Testing the usability of usability testing. *Technical Communications*, September, 343-358.
- Walbridge, S. (2000). Usability testing and libraries: the WSU experience. *ALKI*, 16(3), 23-24.