Usability Testing for Greater Impact: A Primo Case Study

**Joy Marie Perrin,
Melanie Clark,
Esther De-Leon,
Lynne Edgar**

ABSTRACT

This case study focuses on a usability test conducted by four librarians at Texas Tech University (TTU). Eight students were asked to complete a series of tasks using OneSearch, the TTU Libraries' implementation of the Primo discovery tool. Based on the test, the team identified three major usability problems, as well as potential solutions. These problems typify the difficulties patrons face while using library search tools, but they have a variety of simple solutions.

INTRODUCTION

The Texas Tech University Libraries' Usability Taskforce was created to inform, facilitate, and promote usability initiatives for services supporting teaching, learning, and research. The team's first assignment was to study the Libraries' new implementation of Primo, a discovery tool by Ex Libris, which is capable of simultaneously searching all library resources. Primo, branded OneSearch for public use at the TTU Libraries, was initially implemented with no further customization. Library administration charged the team to evaluate the Primo interface as set up, determine whether the tool served patrons in an intuitive way, identify problem areas, and share possible improvements with the Library Systems Group. The issues the team encountered, problems found, and lessons learned along the way are relevant across all library usability efforts and may assist other organizations in developing better searching tools.

The purpose of this study was to evaluate how well OneSearch served library patrons and to identify ways it could be improved before it replaced the existing library search tools. The information collected about the website navigation and searching practices of TTU students was also expected to assist instruction librarians in teaching students how to use OneSearch.

METHOD

The usability study comprised two components to evaluate both OneSearch use and patron thoughts, comments, and observations. The first component was a series of seven

Joy Marie Perrin (joy.m.perrin@ttu.edu), is Assistant Librarian, Digital Resources Unit, **Melanie Clark** (melanie.clark@ttu.edu) is Associate Librarian, Architecture Library, **Esther De-Leon**(esther.de-leon@ttu.ed) is Assistant Librarian for Electronic Resources, **Lynne Edgar** (lynne.edgar@ttu.edu) is Assistant Librarian, Library Systems Office, Texas Tech University Library, Lubbock, Texas.

tasks that participants completed using OneSearch while the team observed. Each participant was guided through the process by a facilitator who would prompt the participant when he or she got stuck. The rest of the team observed both the participant's screen movements and audiovisual footage of the participant's facial reactions from another room, with the help of TechSmith's Morae usability software. While Morae made the observation process easier, the same results could have been achieved through simple screen-capture software, a video camera, or simple note taking by the facilitator. In addition to the observation, the team used Retrospective Recall, asking the patrons to think through their choices after the tasks were done and explain their process.[1] The second component to the study was the System Usability Scale (SUS), a standard survey used to evaluate systems based on self-reported user experience.[2] Participants completed the SUS survey after finishing the tasks.

For the first component, the tasks were developed to cover seven types of materials a patron might find using the search tool:

1. You are looking for a work called "Operations management" by Roberta Russell. Find out if the library has this book and if you can check it out. If the library has it, where is it located?
2. You are not on campus, but you want to read a book about human resources management. See if there are any books available online and if you find them, try to read one of them.
3. Find the database JSTOR and open it.
4. You need to find a full-text online article about customer service training. Try to find an article and view it.
5. You want to see a picture of someone from the 1977 volume of the Texas Tech yearbook (La Ventana). Locate the yearbook and then open the first page of 1977.
6. Find Dr. Rebecca K. Worley's TTU thesis The House of Yes: a Director's Approach and find the abstract of the thesis.
7. You need to find a picture of Frank Lloyd Wright's Fallingwater. Find the picture and access it.

The order of the tasks was important to test the learnability of the system and minimize user frustration. The first task was a simple book search, allowing participants to familiarize themselves with the tool. The difficulty of the searches increased over the next three tasks. The method of "start easy, finish hard" is recommended in the CUEP workshop to test memorability and learnability of the website.[3] However, the team varied this model by designing the last three tasks to be similar to the first two. These tasks each requested a different material type, but all the materials could be found with a simple search identical to that of the first task. This design showed whether participants learned how to use the system and remembered the process. Participants struggling with the last tasks would indicate a severe usability problem.

The team timed participants' performances with each task and noted each error or problem encountered. Each task was labeled as either a success or failure for each participant depending on whether he or she completed the task, had to be guided to complete it, or gave up.

PARTICIPANTS

Eight patrons participated in the study. From a demographic profile that participants filled out prior to completing the tasks, the team identified three expert users, three intermediate users, and two novices. Student classification, how frequently the participant used the library website, and the ways the participant used the library website all factored into the user status.

RESULTS

System Usability Scale Score

The System Usability Scale "grades" a website or system by how usable patrons perceive the system to be, resulting in a single number score. A score above 80.3 is in the top 10 percentile and is considered to be an A. A score of 68 is average, with anything under 68 below average.[4]

OneSearch received a SUS score of 78.25 from the eight participants of the study. This is comparable to a B+, indicating that overall, the implementation of OneSearch was successful, at least in terms of how students perceived it after using it. To identify specific problems with the interface, the team looked at three factors of the participants' performances on the tasks.

Average Time Spent on Each Task

Figure 1 shows the statistics of the seven tasks. As expected according to the average completion time, participants spent more time on the first task. It may be inferred that they were acquainting themselves with the system. The fourth task, which the team expected to be the most difficult, proved to have the longest completion time. From the completion time alone, the team was unable to determine if task 4 was problematic or not. While participants started their search with the OneSearch interface, they had to wait for a separate integrated system, such as a citation linker, to retrieve any found articles. This added to the task completion time.

Task	Average Completion Time (min)	Completed with Ease	Completed with Difficulty	Average Error Rate
1. Find a book	1.74	50%	50%	1
2. Find an e-book	0.97	87.5%	12.5%	0.63
3. Open a database	1.11	37.5%	62.5%	3.25
4. Find an article	2.92	37.5%	50% (1 did not complete the task)	3.63
5. Find a digital collection item	1.12	62.5%	37.5%	0.75
6. Find a thesis	0.87	87.5%	12.5%	0.38
7. Find an image	0.70	87.5%	12.5%	0.38

Table 1. Task Results

A more telling observation was that the time on tasks, excluding task 4, diminished between task 1 and task 7, suggesting that participants had no trouble learning and remembering how to use the system.

Error Rate for Each Task

The error rate proved to be the most accurate indicator of usability problems with each task. Each time a participant chose a wrong path, faced an impasse, or had to be guided by the facilitator, the event was labeled as an error. In the ideal scenario, the participants would make no errors; therefore, the more errors observed, the greater indication of a usability problem.

As seen in table 1, although the first task took an average of 1.74 minutes, the eight participants tended to only make a mistake once while trying to find a book. Tasks 2, 5, 6, and 7 had an average error rate below 1. Because the error rate seems to decline from task 1 to task 7 (excluding tasks 3 and 4), the team inferred that users were able to learn how to use the system quite easily.

Tasks 3 and 4 seemed to cause problems. Both of these tasks had an average error rate above 3. This indicated to the team that, since the SUS score for the entire system was good, they needed to identify why the database search and article search were problematic.

Success Rate for Each Task

Table 1 also shows the three ways that tasks were tagged during the study. The participants completed the task with ease, completed the task with difficulty, or failed to complete the task. As the team expected, the first task was divided between those who completed it with ease and those who completed it with difficulty. This was expected when the participants were learning the system. Task 2 shows a marked improvement: 82.5 percent of the participants found an e-book with ease. Tasks 3 and 4 however, show up again as problematic. One of the participants failed to complete task 4. After tasks 3 and 4, the participants successfully regrouped, with 87.5 percent completing the last two tasks with ease.

Average Completion Time							
User Level	**Task 1**	**Task 2**	**Task 3**	**Task 4**	**Task 5**	**Task 6**	**Task 7**
Novice	1.41	0.62	0.93	2.35	0.57	0.72	0.67
Intermediate	2.28	1.13	1.57	1.81	1.27	0.86	0.37
Expert	1.44	1.04	0.77	4.42	1.35	0.98	1.06

Average Error Rate							
User Level	**Task 1**	**Task 2**	**Task 3**	**Task 4**	**Task 5**	**Task 6**	**Task 7**
Novice	0	0	3.5	4.5	0.5	1	0.5
Intermediate	2.33	0.33	3.67	1.67	0.67	0.33	0.33
Expert	0.33	1.33	2.67	5	1	0	0.33

Table 2. Task Results by Experience Level

Results by Experience Level

Based on the team's observation, novices used the simplest approach to each task. The intermediate and expert users sometimes extended their task completion time by performing more complex searches. Almost without exception, the first thing the more experienced users did for each task was to look at the dropdown menus or go to "advanced search," even when a general search would provide sufficient results. As shown in table 2, this approach lengthened their completion time and increased their error rates on the first two tasks, but the novice error rates increased dramatically on the most difficult tasks (3 and 4), surpassing those of the intermediate and expert users. The tendency of more experienced users to perform complex searches did not negatively affect the their success in completing each task.
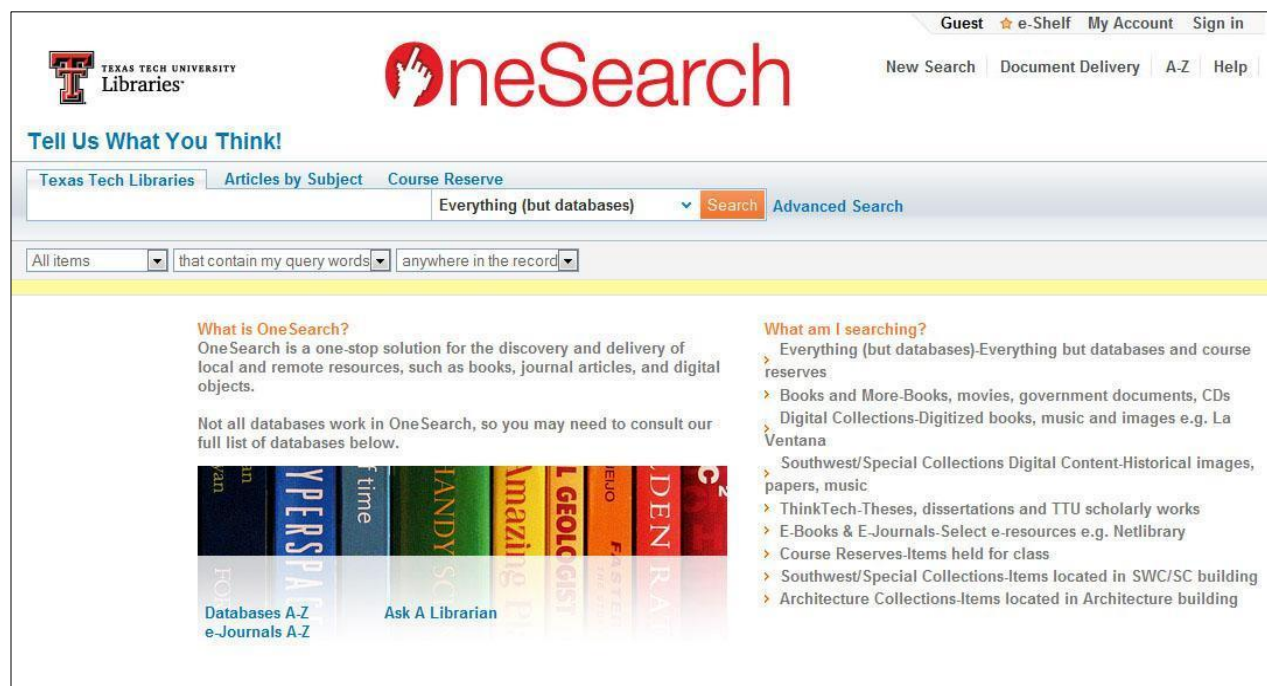

Figure 1. OneSearch Home Page

Problems Identified

Figure 1 is a screen shot of the OneSearch interface used for the study. As shown, there are three tabs for different searches. The first tab, "Texas Tech Libraries," searches the print collections, institutional repository, digital collections, special collections, and e-books. The second tab, "Articles by Subject," is a federated database search that can be narrowed by subject. To find a specific database, users had to click on the A–Z link either in the upper right corner or at the lower left under the image.

The test indicated that users faced the most trouble when searching for articles and opening specific databases. The team identified three problems based on the difficulties the participants experienced. All of these problems were related to visual design and clarity rather than the basic functionality of OneSearch. Identifying problems, however, does not address how to best resolve those problems. Observing participants using the interface, as well as researching current web standards, may lead to educated guesses as to possible solutions. Knowing the limitations of how the OneSearch interface design could be changed, the team decided to offer as many possible solutions as could be identified. This was to highlight that there was no single way to fix the problems, but many different ways the problem could be solved. The team put the options in order of their expected effectiveness on the basis of what was observed during the test.

Problem 1: Individual Databases are Difficult to Find

While analyzing the task completion statistics and video footage, it became clear that the way to access the databases was not visible to users. As shown in figure 1, the "Databases A–Z" link was at the very bottom of the page in a location not immediately noticeable. In addition, the formatting on the link did not look like an obvious link. Even if users saw it, they might not realize it was something they could click on.

Solution 1: If possible, make "Find Databases" a search scope in the dropdown menu. The participants were more willing to go to the menu than to look around the page. This meant that the best place for users to be able to search Databases was by selecting them on the dropdown menu. This solution was not possible in OneSearch, however.

Solution 2: Make "Find Databases" a fourth tab. Participants were also more likely to look through the tabs than they were to see the database link at the bottom of the page.

Solution 3: Move "Find Databases" or "Databases A–Z" to a different part of the page. Figure 2 is a mock-up of different ways this option could be implemented. This was to highlight that as long as the link was put higher on the page, in a more visible place, users would be more likely to find it.

Solution 4: Make "Databases A–Z" bigger or more eye-catching by changing the color. This would increase the likelihood of it being seen.
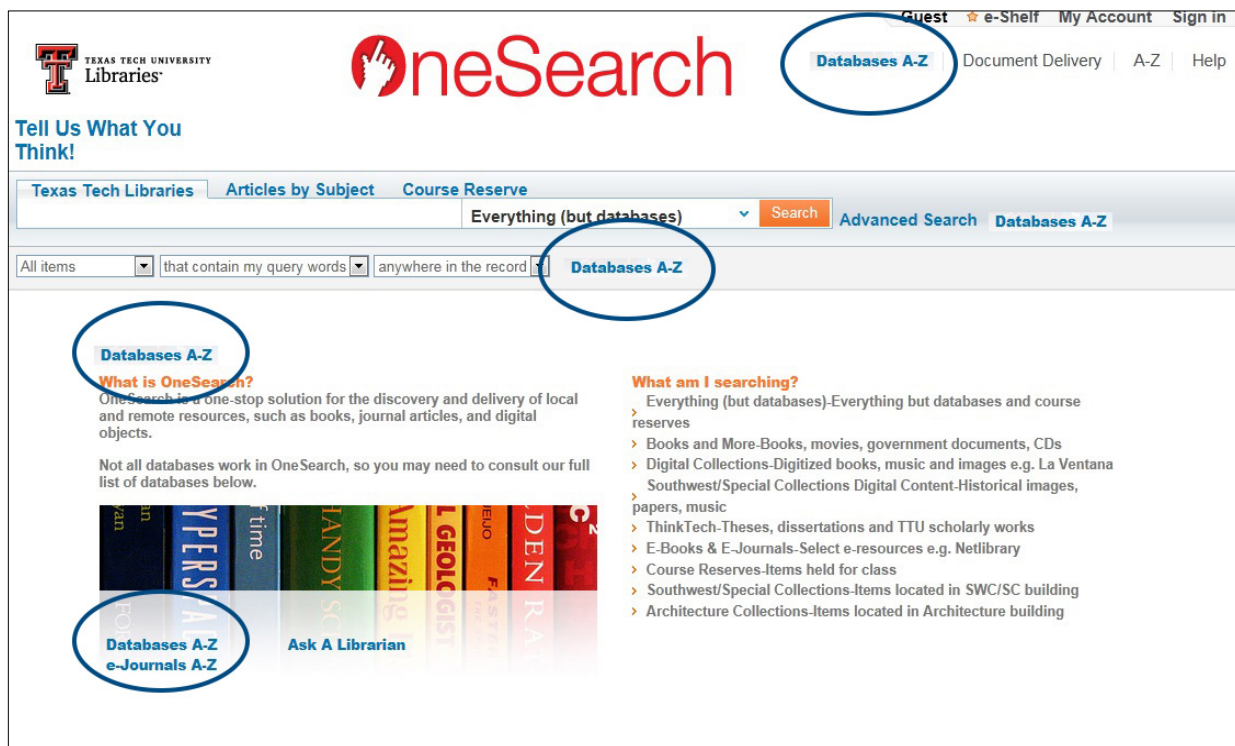
Figure 2. Suggested Locations for Databases A–Z

Problem 2: Dropdown Limiters are Misleading

In figure 1, three search limiters can be seen below the search box. The ones shown in figure 1 are "All items," "that contain my query words," and "anywhere in the record." The first box provides a way to limit the results by type such as "Books" or "eJournals". The second box offers a choice between a general search for query words, specifying if the search is an exact phrase search or if a field starts with the query words. The last box offers a way to specify which field the query is searching. Some of the participants ran unsuccessful searches during the test, and then erroneously tried to get results by "limiting" the items in their already faulty search. For example, a few students went to the "Articles by Subject" tab and searched for an image (task 7). When no images came up, they went to the limiters and chose the format "images." This resulted in zero results returned because there were no images in the "Articles by Subject" tab. Again, the team gave three different options in order of their perceived effectiveness.

Solution 1: On the "Texas Tech Libraries" tab, remove "articles" from the dropdown limiter. The reason behind this was that there were no articles in the basic search, so it should not be an option.

Solution 2: On the "Texas Tech Libraries" tab, remove "databases" from the dropdown limiter (unless "Find Databases" can work as a search scope).

Solution 3: On the "Articles by Subject" tab, remove "images" and "journals" from the dropdown limiter.

However, most participants did expect all of the content to be in the main dropdown menu and did not tend to use the tabs or limiters. The team offered a mock-up of the best possible scenario in which the only options to search were in the main dropdown (figure 3). This would have been the most intuitive way for patrons to search, but it was also the most technically complex to implement.
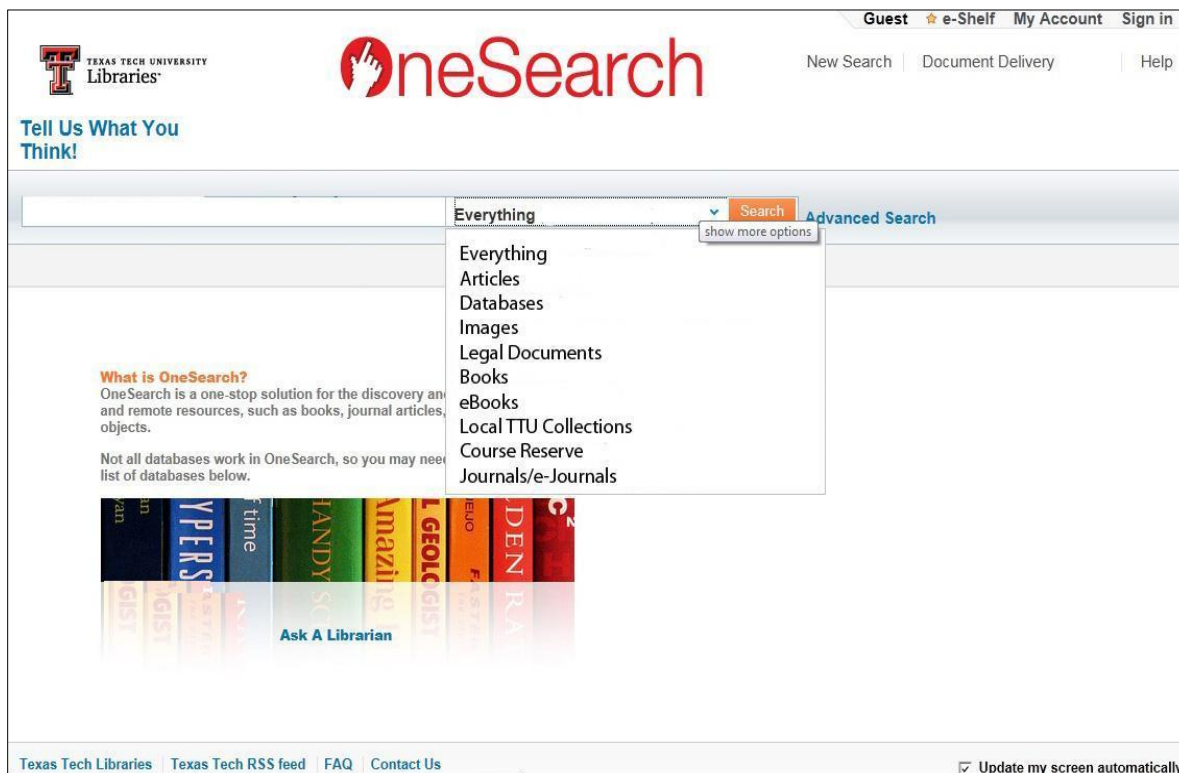

Figure 3. Dropdown Menu Suggestion

Problem 3: "Articles by Subject" Tab is not Visible to Users

The "Articles by Subject" tab, which allows users to choose a federated database search by subject, was not visible to participants. The text was smaller than other surrounding text and not immediately recognizable as a tab. One of the reasons that the tabs were removed from figure 3 is that they were not easy for users to see. The team recommended three options.

Solution 1: Make the tabs more visually identifiable by designing them more like traditional website tabs.

Solution 2: Enlarge the tab text so that the tabs are more visible. Figure 4 shows a mock-up with enlarged, more noticeable text.

Solution 3: Add an "Articles by Subject" description to the "What am I searching?" text. The team noticed that the explanatory text on the right side of the page did not include a description of the "Articles by Subject" tab.
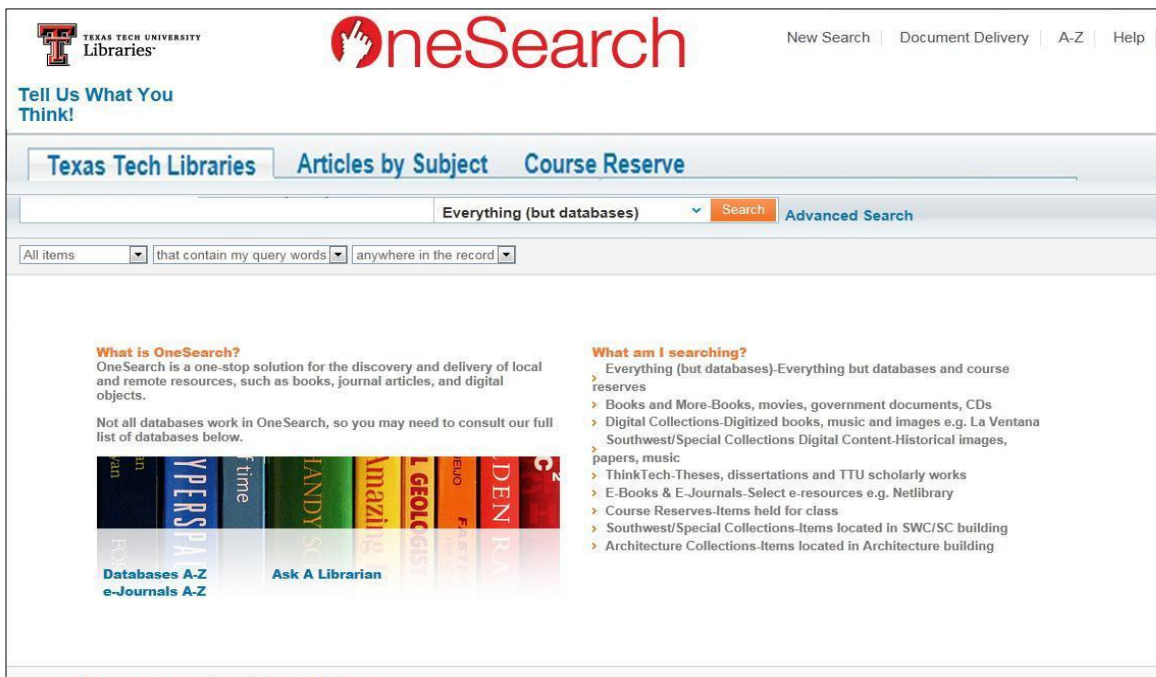
Figure 4. Enlarged Tabs

Other Minor Problems

In addition to the three major problems, the team noted smaller issues. One of the problems was that some of the titles of the dropdown search scopes were not in terminology that the students understood. For example, one of the scopes is "ThinkTech," the name of Texas Tech's institutional repository. Since this name doesn't indicate what the scope actually searches—mainly theses and dissertations—users didn't know what was in "ThinkTech" unless they read the explanatory text on the right. The team recommended changing the scope name to something more descriptive, such as "Theses, Dissertations, Faculty Research."

Another small usability problem was how difficult it was to see the indication of an incorrect search. The "Did you mean?" spelling suggestion on the search result page was very small, smaller than the notification that there were no results. Participants who made simple spelling errors didn't realize they had failed because of this simple error and assumed there were no results.

DISCUSSION

The team submitted these findings to the Library Systems Group in two different ways: a written report and a presentation including the above mock-ups, charts, and video footage from the usability study. The video clips allowed the team to both illustrate the problems and show the Systems Group the sources of those problems. The presentation with more multimedia content made a much greater impact than the written report and resulted in the Systems Group better understanding the problems and how to fix them. Visual aids are an effective way to report results because of how crucial visibility is in usability. One example of how the presentation was more

effective is that the team played a game with the Systems Group by showing figure 2 without the "Databases A–Z" links circled. The Systems Group was asked to raise their hands when they first saw the link. After everyone raised their hands, the next slide showed all the locations that the "Databases A–Z" link could be found. Most of the group had not realized there was more than one link, illustrating that some positions are more noticeable than others and that users tend not to linger on a page. This was more effective than a dry report stating this fact.

Implementing usability findings is often more difficult than identifying them, particularly when usability is conducted on a "finished" system. Not all the problems the team identified were able to be addressed, but some problems were fixed quickly and easily, such as including the link "Find Databases" at the top of the page (see problem 1, solution 3 above). What the usability study did was allow the Systems Group to understand how patrons view their tool and how they are likely to work with it.

CONCLUSION

These small changes made the system more usable for patrons, which is what usability testing is all about. It is less about making a system conform to a single way of doing things than finding small ways that the system can be made easier to use. Changing the name of the search scopes or changing the position of a link is a relatively small investment of time and resources that yields great benefits for patrons in making the system easier to use.

One of the most interesting observations from this study was that most of the users wanted all their search options in one place. They preferred one dropdown menu to handle all their needs. For future development on these kinds of systems, this kind of preference should be kept in mind. The majority of patrons might be happier with a tool with less capability and simpler options rather than a complex tool with many different ways to approach their search.

REFERENCES

1. Brian Still and M. **[QY: first name?]** Betz, *A Study Guide for the Certified User Experience Professional (CUEP) Workshop* (Lubbock: Texas Tech University, 2011), 61.

2. Jeff Sauro, "Measuring Usability with the System Usability Scale (SUS)," Measuring Usability, February 2, 2011, http://www.measuringusability.com/sus.php.

3. Still and Betz, *A Study Guide,* 67.

4. Sauro, "Measuring Usability with the System Usability Scale (SUS)."