

Discovering Usability: Comparing Two Discovery Systems at One Academic Library

Mireille Djenno, mdjenno@uic.edu
Glenda Insua, ginsua1@uic.edu
Gwen M. Gregory, ggregory@uic.edu
University of Illinois at Chicago, Chicago, Illinois, USA

John S. Brantley, jsbrantley@eiu.edu
Eastern Illinois University, Charleston, Illinois, USA

Re-submitted June 2, 2014

ABSTRACT

In the spring of 2013, the University Library at the University of Illinois at Chicago was in the unique position of having access to two discovery systems, Summon and WorldCat Local, at the same time. When tasked with choosing between the two systems, librarians undertook a usability study of Summon and WorldCat Local. The goal of this study was two-fold: to test the ease-of-use of each discovery system with an eye toward identifying one tool to retain for the longer term, and to learn about the search behaviors of different types of user groups. Eighteen subjects, consisting of undergraduate students, graduate students and faculty, participated in the study. Participants performed usability tasks using each tool and answered pre-task and post-task questions. While there was no clear preference among study participants for either discovery layer, individual groups did express preferences. Faculty, for example, preferred Summon to WorldCat Local at a rate of five to one. The study findings are explored in detail through an examination of the three major data sets produced by the usability test instrument: results derived from tasks performed by participants as part of the study; themes and trends identified by the investigators within the recorded participant tests; and discovery tool preferences as determined from pre-task and post-task questionnaires administered to study participants. This study has implications for librarians engaged in information literacy instruction; those considering implementing discovery tools, as well as for librarians currently using Summon or WorldCat Local at their libraries.

KEYWORDS usability testing, discovery systems, discovery tools, user preferences, students, faculty, one search, information literacy, WorldCat Local, Summon

INTRODUCTION

In the continual quest to provide a simple search interface for the information available in multiple formats through academic library Web sites, many libraries have implemented the latest generation of search products, known as “discovery tools.” These products allow users to search across multiple databases at the same time and help the searcher “discover” collections and resources that are not described by the indexes of large commercial library databases and are therefore hidden from view. These hidden collections are available only through the local institution’s search interfaces and remain marginalized on subject-specific, low-traffic library Web pages.

Discovery tools, also known as “Web-scale discovery,” allow the library to provide access to multiple collections through a single search interface, from the catalog and journal holdings to local digital collections and institutional repositories. Two of these tools, Summon and WorldCat Local, the discovery products of Serials Solutions and OCLC respectively, were simultaneously available to the library users of the University of Illinois at Chicago (UIC) in the spring and summer of 2013.

Summon, like most discovery tools, presents the library user with a single search box interface that promises to make all of a library’s content accessible by assimilating the bibliographic data from the library catalog, other local collections, information harvested directly from large journal publishers as well as selected abstracting and indexing metadata from commercial databases, into its large central index. This creates a search tool that provides books, articles, archives, and digital collections in one set of search results.

WorldCat Local is OCLC’s solution to the same problem. WorldCat Local’s central index has the largest catalog of bibliographic records in the world. To that index, WorldCat

Local adds the article-level records for many additional databases owned or licensed by OCLC. In both cases, the library client of either tool can access content to which they already subscribe, and then add records for local collections.

It is unusual for an institution to have simultaneous access to two discovery tools as they are competing entities that perform primarily the same functions, and because maintaining two discovery systems is cost prohibitive for most universities. UIC was an early adopter and beta tester of the WorldCat Local discovery tool. UIC was also an early customer of Summon in Illinois, and as a member of a large consortial catalog – the I-Share catalog of the Consortium of Academic and Research Libraries in Illinois (CARLI) - UIC provided Summon developers the opportunity to identify issues and problems surrounding the display of multiple item locations and local holdings versus consortial borrowing access (see Brantley, Gregory, Lambrecht, Wang, “Summon in the I-Share environment” CARLI Discovery Layer Webinar Series 2013). Problems arising from library tests of Summon in UIC’s consortial environment delayed UIC’s launch of Summon as its default discovery tool and necessitated the extension of the WorldCat Local license another year. The result of these circumstances was simultaneous access to two discovery tools and the opportunity to draw direct comparisons between the usability of each, as well as between the search experiences of different user groups.

LITERATURE REVIEW

Since 2007, when WorldCat Local entered the market, several additional options have become available: Serials Solutions’ Summon, EBSCO Discovery Service, Ex Libris Primo Central, and Innovative Interfaces Encore Synergy (Vaughn 2011). Several usability studies

have been conducted on single discovery systems, or studies of multiple systems with each system at a different institution. Although libraries continue to make important and costly decisions in choosing from among these discovery systems, as of this writing, the investigators have found no studies comparing two of the most-widely adopted discovery tools, Summon and WorldCat Local, to each other within the same library. Melissa Becher and Kari Schmidt (2011) conducted a comparative study of WorldCat Local and Aquabrowser at the same library; however Aquabrowser is not technically a discovery system, but instead a catalog overlay tool providing sophisticated search and sort facets to a library catalog. It does not include the central index element that allows users to search across multiple databases and collections.

A number of libraries have conducted usability studies of Summon to learn how users employ the tool in searching for information. Julia Gross and Lutie Sheridan (2011) conducted a usability study of Summon with first-year students in computer science and library and information science. Students were asked to complete four typical library tasks, while researchers observed and recorded movements with Techsmith's Camtasia software (<http://www.techsmith.com/camtasia.html>). Their findings showed students "were confident with the user interface, but somewhat perplexed by the search results" (242). Students liked the simplicity of having a single search box, but found it difficult to interpret their results and to differentiate between formats. Similarly, Anita Foster and Jean MacDonald's (2013) study of Summon at Illinois State University demonstrated that users were unable to differentiate between source types in their results list and found it difficult to determine if an item was physically in the library.

Andrew Asher, Lynda Duke, and Suzanne Wilson's (2013) study of Summon, Ebsco Discovery Service (EDS), and Google Scholar went one step further and found that regardless of the tool being used, students had a difficult time evaluating sources, often settling for less-than-adequate resources simply because they appeared on the first results screen. Overwhelmed by the number of search results, students relied on the tool for evaluation, "using the search engine's relevancy rankings to determine relative quality" (474). As relevancy ranking algorithms are proprietary, and therefore unknown to the user, this reliance is problematic. For instance, Summon seems to rank newspaper articles fairly highly according to a number of studies (Asher et al. 2013; Way 2010), even though students often need more scholarly resources for their research. The way libraries configure these tools can also factor into rankings. These findings stress the need for information literacy instruction and "underscor[e] the continued need for research training" on discovery tools of all types (Asher et al. 2013, 476).

Although discovery tools may seem intuitive, at the University of Colorado at Boulder, Rice Majors (2012) discovered that students need instruction to use and search them effectively. He sought to test the "assumption of ease of use" inherent with discovery systems by conducting a comparative study on five different tools: Encore Synergy, Summon, WorldCat Local, Primo Central, and EDS (186). His study showed that students experienced difficulties completing tasks across all systems, particularly in interpreting search results. Interestingly, students suggested "enrichment options" to make searching and evaluating results similar to Amazon's online shopping experience - including options such as ratings, reviews, lists of similar items, ability to search within a book, and the like (196). While Summon has a feature it calls a "database recommender" when a user's results are coming

primarily from publications also indexed by a subject specific database, discovery systems generally do not yet provide the level of sophistication suggested by the students in Majors' study.

Majors' conclusion that users find it difficult to evaluate results using WorldCat Local were echoed by Sue Fahey, Shannon Gordon, and Crystal Rose's (2011) usability study comparing WorldCat Local with their library catalog. They tested two user groups, students in the arts and sciences and health sciences students and faculty, and neither group expressed a preference for WorldCat Local over the library catalog. Both groups of users also faced challenges differentiating between formats and various editions of books, although they considered the tasks to be straightforward. This discrepancy between perceived success and actual success is problematic for users and librarians alike.

Conversely, in Becher and Schmidt's (2011) study of WorldCat Local and Aquabrowser, differences in preference did occur between user populations - undergraduate students preferred WorldCat Local, while graduate students preferred Aquabrowser. Researchers could not definitively determine why academic status influenced preference but posited that it had more to do with interface than with content (211). Despite this difference, both groups had some commonalities. Users in all academic groups liked having links to full-text articles, having a results list comprised of both books and articles, and date and format facets (199).

Users in Bertot et al.'s (2012) study also had a fairly favorable impression of WorldCat Local, and as in Becher and Schmidt's study, design and interface were rated highly. Unfortunately, challenges arose when users looked at search results, perhaps not a surprising finding as discovery tools contain many more types of formats than traditional

catalogs. As in studies of Summon (Asher et al. 2012; Foster and MacDonald 2013; Gross and Sheridan 2011), users had difficulties locating items, determining who owned an item, and dealing with multiple or similar results.

The current study builds upon the literature by testing two different discovery tools at the same library, thereby minimizing the differences that arise from comparing systems across different collections and institutional customizations. The current study also looks at a wide range of patron populations and skill levels, from first-year students to faculty.

METHODOLOGY

This study was designed to test the usability of Summon and WorldCat Local, as implemented at the UIC Library, as well as to analyze the search behavior of several different user groups -- undergraduate students, graduate students, and faculty. The usability study was conducted using Barnum's *Usability Testing Essentials* (2011) as a guide. After developing a protocol and receiving approval from the UIC Office for the Protection of Research Subjects, the researchers solicited test participants through e-mails to listservs, flyers posted on campus, and personal contacts. Participants were selected on a first-come, first-served basis. All participants were students or faculty of UIC. Six participants from each user group (18 participants total) were tested. Students were rewarded with a \$25 Amazon gift card. Participants represented a wide range of disciplines, from theater in the humanities to medicine in the health sciences, which allowed the researchers to examine the search behaviors of users in many fields.

The test instrument was divided into four components (see Appendix A). Participants were first asked six questions about their online research, including which online tools they

used most often, and whether they were familiar with Summon or WorldCat Local. Next, the instrument included six tasks to be completed by the participant using first one tool and then the other. The tasks included known-item searches for both books and journal articles, as well as general subject research on an assigned topic and on a topic of their choice. Some of the known items were available at the UIC Library, some through the I-Share consortium, and some would have to be requested through interlibrary loan. Since similar tasks were completed using both tools, it is possible that a small amount of “learning” could occur between using one tool and then another, skewing results in favor of the tool used second. To avoid this, half of the participants were asked to use Summon first, followed by WorldCat Local; half began the test with WorldCat Local, followed by Summon.

Third, after completing the tasks for a tool, participants were asked to reflect on their experiences with that tool and comment on what they liked and disliked. Finally, after completing the six tasks using both tools, participants were asked a series of questions prompting them to compare their experiences with the tools and indicating which of the two tools they preferred. The interviews were conducted with two interviewers present, one asking questions and one taking notes. Eight librarians took part in the study as interviewers. Interviewers read the task questions from a prepared script to ensure consistency. Participants were instructed to talk through their actions as they completed the tasks, and the interview audio and screen activity were captured for later review and analysis using Techsmith’s Camtasia software. Participants were also given a copy of the usability tasks. This simplified known-item searching and reduced the likelihood of misinterpretation of the directions, because participants could read the exact citation and spelling. One hour was allotted for each test, with the majority taking 30-50 minutes to complete.

Qualitative and quantitative data were gathered and analyzed using a combination of top-down (study results) and bottom-up (affinity matching) strategies (Barnum 2011).

Qualitative data were based on usability test tasks and post-test questionnaires and were analyzed using an affinity matching exercise wherein investigators reviewed the recordings of the usability tests and made notes of recurring actions, themes and trends (e.g., electing to use the advanced search function) within the recorded interviews. These notes were then grouped by similarity of action, theme, or trend. Using a modified version of the constant comparative method (Glaser and Strauss 1967; Gorman and Clayton 2005) used by Becher and Schmidt (2011), the researchers gathered the data from notes taken during the usability tests and notes from the affinity matching exercise. These data were then coded into three primary categories: functionality, content, and interface.

Quantitative data were based on the tasks performed in the usability tests. Benchmarks were established for known-item searching and retrieval tasks performed as part of the study using the following method:

- 1) Tasks with objectively measurable “success” or “failure” (e.g., finding a specific title using the discovery tool) were selected as indicators.
- 2) The correct answers (or range of possible correct answers) and incorrect (or range of possible incorrect answers) to these questions were incorporated into a new recording sheet that the investigators used when listening to recordings/screencasts of the interviews. Task success was denoted by “S” and task failure by “F.”
- 3) Performance on each task was coded simply with a single point assigned for each “success” (S) and no points assigned for “failure” (F) to satisfactorily complete a task.

To assure inter-rater reliability, for each of the 18 tests, task “success” and “failure” was determined by two investigators.

- 4) These data were then analyzed within four categories: availability, retrieval, filtering, and e-mail, which allowed the researchers to ascertain the usability of the tools based on key aspects of the user experience.

RESULTS

Analysis of the usability tests revealed that the most frequently used WorldCat Local function was the advanced search, used by 44 percent of participants, followed by the filtering functionality, used by 33 percent of participants. Many participants were confused by the e-mail function and indicated that they would prefer to bypass it altogether. Twenty-two percent of participants changed their initial search terms when doing a subject search in WorldCat Local, rather than using the filtering functions to narrow their results.

Functionality results measured how successful users were when using the tools to determine the availability of items within local or consortial holdings, facilitate the retrieval of items, filter results, and e-mail select results (see Appendix B for detailed scoring). For WorldCat Local, the overall availability score was 93 percent – meaning that users could determine the availability of the item they were looking for in 93 percent of the cases when searching with WorldCat Local; the mean book availability score was 91 percent – meaning that when searching for *books* specifically, users were able to determine the availability of the book 91 percent of the time, and for online article availability, the mean WorldCat Local score was 100 percent. The mean score for book retrieval tasks for WorldCat Local was 53 percent.

The mean score for WorldCat Local filtering/limiting tasks was 77 percent. The mean score for WorldCat Local e-mail functionality was 67 percent.

Two participants noted that they liked the layout of WorldCat Local and four participants found it easy to determine the location of an item (see Table 1).

Beyond a preference for full-text resources, only one participant mentioned the content of the articles they found while testing WorldCat Local.

[PLACE TABLE 1 HERE]

Users displayed similar behaviors when using Summon. Eleven participants elected to use facets to filter search results when testing Summon; seven participants used the advanced search function; six participants used the date limiter to filter their results and two used the full-text limiter. Three of the participants used alternative search terms rather than using facets to refine their search.

The mean score for the book availability questions for Summon was 81 percent (see Appendix B). The mean score for book retrieval tasks for Summon was 56 percent. The mean score for Summon filtering/limiting tasks was 83 percent and the mean score for Summon e-mail functionality was 67 percent.

Six out of eighteen, or 33 percent, of Summon users did not scroll down far enough on the screen to determine the availability of items. For five participants, the UIC holdings were not obvious. For six participants, the location of the item they were looking for was unclear. Three participants did note that they liked the date limiter (see Table 2).

As with WorldCat Local, only one participant mentioned the content of articles found while testing Summon.

[PLACE TABLE 2 HERE]

DISCUSSION

Comparing the tools - Preferences

Participants were evenly split between those who preferred Summon and those who preferred WorldCat Local, although interestingly, participants overwhelmingly considered Summon the easier tool to use. Fifty percent of participants found Summon easiest to use, versus 28 percent for WorldCat Local, and 22 percent having no preference.

Most participants said they would use both of the tools again (10/18 or 56 percent), but among those not wanting to use either tool again, two graduate students cited their preference for using a single database or journal for their searches. Those identifying just one tool they would use again chose Summon over WorldCat Local at a rate of four to one.

More than half of the participants chose Summon when asked which tool should be retained in the hypothetical scenario that only one could be kept. The remainder was split between 11 percent who had no preference and 33 percent who stated a preference for keeping WorldCat Local.

While the vast majority of participants had not used either tool before, the researchers assume a degree of comfort with Web-based tools for a few reasons. Sixty-seven percent of participants rated their Internet searching skills good or excellent, and five out of twelve participants had taken a library research class at some point. (Faculty members were not asked

if they had taken a library research class.) Most participants also reported using an online tool on a daily basis (Google being that tool for 61 percent of participants). Participants were not provided with a definition of online tools for this open-ended question, but rather were provided examples of online tools that were a mix of online research databases, e.g., JSTOR, PsycINFO, and Web-scale discovery systems such as Summon. Because participants were not asked to articulate their understanding of the term “online tool,” it is impossible to say with any certainty whether or not they were making meaningful distinctions between tool types when answering the question. However, there seemed to be no statistically significant advantage for those who had used the tools before.

Comparing the groups - Preferences

Notably, most of those preferring Summon were faculty members (56 percent) with the rest of the groups preferring Summon equally at a rate of 22 percent each (see Fig.1). Only one faculty participant preferred WorldCat Local and within this user group, there were no respondents who indicated “no preference.” Faculty therefore preferred Summon at a rate of five to one over WorldCat Local, the strongest indicated preference among all of the study data. Among their reasons for preferring Summon, faculty indicated that fewer steps were needed to get results; they liked the facets and Summon provided what they considered better results. This preference rationale aligns with the three critical measures of usability outlined by Carol M. Barnum: effectiveness, efficiency and satisfaction. Barnum notes that “*effectiveness* and *efficiency* support the user’s need to achieve a goal for using the product with accuracy and speed” (Barnum11).

Undergraduate and graduate students preferred WorldCat Local overall, (see Fig.1) and did so at a rate of two to one over Summon. Faculty members voted almost unanimously for retaining Summon, which echoed their overall discovery tool preference. Graduate students were also consistent in voting to retain the tool that they preferred. However, undergraduates preferred WorldCat Local two to one (four preferred WCL and two preferred Summon) but were split about which tool to retain: only two users voted to keep WorldCat Local, three voted for Summon, and one had no preference. It is interesting to note that undergraduates were the only group for whom Summon proved more usable, as measured by their performance on the usability test tasks. In other words, despite their stated preference for WorldCat Local, they were more successful in completing the usability tasks when using Summon.

[PLACE FIGURE 1 HERE]

Comparing the tools - Usability

As with WorldCat Local, one third of participants preferred to bypass the built-in e-mail function when using Summon, a choice that surprised the researchers as this is now a standard (and presumably desirable) feature of Web-scale discovery tools and databases. However, users did use the available facets to a large extent. Sixty-one percent of Summon users (as opposed to 33 percent of WorldCat Local users) used the filters, which might suggest that filters were easier to use in Summon. The mean score for success with filtering or limiting tasks in Summon was 83 percent (see Appendix B). This is the first of two functions in which Summon outperformed WorldCat Local, the second being book retrieval.

Summon and WorldCat Local's overall usability scores (75 percent and 81 percent, respectively) were very similar to their usability scores for known-item book searching tasks (70 percent and 78 percent, respectively) (see Appendix B). However, the mean score of 81 percent for the book availability questions for Summon was much less than the mean score for book availability (91 percent) in WorldCat Local. The mean score for book retrieval tasks for Summon (56 percent) was very close to the retrieval tasks for WorldCat Local (53 percent), an indication that regardless of the tool used participants found it challenging to retrieve books once they had determined their availability and location. Twenty five percent of users had difficulty determining the availability of items when using Summon. This seemed to strongly correlate with not scrolling down far enough on the results page to determine if the item was available. Of equal concern (and less easily accounted for), was the difficulty in discerning local holdings (21 percent) and item locations (25 percent), experienced by a significant number of participants when testing Summon. These findings were similar to those of Foster and MacDonald (2013) who also describe users having a difficult time determining the availability of search results when using Summon.

Similar to the outcome of WorldCat Local testing, only one user mentioned the content of the articles found during the Summon usability tests. This finding is of note as the post-test questions indicated that when asked which discovery tool provided the best results, 50 percent of participants said Summon; 28 percent said WorldCat Local and the remaining 22 percent stated they had no preference between the two tools. The top reasons they cited for considering the results from Summon better were that they were more "comprehensive" and "more relevant." The highest proportion of participants indicated that relevance of search results was the basis of their preference. Interestingly, this criterion for tool preference was

equally represented among the groups; at least one user from each group invoked “more relevant” when asserting that Summon provided better results.

Whereas the responses to the question “Which tool would you keep if you could only keep one?” aligned with the results of overall tool preference, there was a much greater discrepancy between the tool participants would choose to keep and the tool they felt provided the best results. In other words, all groups stated that relevance of the results was the basis of choosing Summon’s results over those of WorldCat Local, whereas not all groups would elect to retain Summon if the library were to keep just one tool.

An overall availability success score of 89 percent for all participants indicates that participants were able to easily determine the availability of the item they were tasked with finding, regardless of the tool being used and regardless of the item (see Appendix B). Therefore, although WorldCat Local proved to be the more usable tool, both tools met the usability measures of efficiency and effectiveness for a central function, namely allowing users to determine the availability of an item within the library’s holdings.

Comparing the groups - Usability

Although its findings are not statistically generalizable, this study was nonetheless designed with particular attention paid to creating a group of study participants representative of a cross-section of users. For this reason, a closer examination of the usability results by user group, and their ranking relative to one another (see Table 3) is warranted.

Graduate students had the highest total rate of successful completion (82 percent), while the undergraduate group was least successful overall (70 percent) in completing the

usability tasks. However, overall success rankings were not echoed in more granular examinations of the data categories.

Within the overall availability category (the ability for users to determine the availability of an item), the graduate student group fared best (91 percent), followed by the undergraduate group (89 percent), and the faculty group (82 percent). On tasks specifically dealing with determining the availability of a book, graduate students ranked the highest (89 percent), followed by undergraduate students (85 percent), and then faculty (80 percent). In online article availability, the undergraduate students were most successful (100 percent), followed by the graduate students (96 percent), and then faculty (87 percent). Taken together, these data show that although the undergraduate group scored the lowest in overall task successes, it was the faculty group that had the most difficulty determining the availability of items using the tools.

In the retrieval category, graduate students were most successful (67 percent), followed by faculty (57 percent), and undergraduate students (39 percent). The finding for this category echoes the overall successes across groups.

For the limiting tasks, which asked participants to limit results to a year range and to a shorter list they considered more relevant to their understanding of their chosen topic, the graduate students ranked first (85 percent), faculty members second (80 percent) and undergraduate students third (62 percent).

In the category devoted to distinguishing between material types, faculty ranked first (87 percent) with graduate students following (79 percent) and undergraduates coming in last (62 percent). This echoes the findings of the Gross and Sheridan (2011) study of first-year undergraduate students who had difficulty discerning differences between material types.

In the ability to use the e-mail function, graduate students were most successful (79 percent), the faculty followed (63 percent) and by a very slim margin, the undergraduate group came last (62 percent). This is another ranking that mirrored the overall task success rankings.

There was significant variability among the groups in their relative rankings (see Table 3). Although the faculty weren't the lowest in overall successes, upon examination of sub-categories, their ranking among the group was, with one exception (material type), either second or third. It is difficult to speculate why this would be. Anecdotal observations suggest that faculty members are less flexible searchers, having used their favorite research tools and methods for a long time. It might also be significant that the faculty group was the only group that did not include digital natives.

[PLACE TABLE 3 HERE]

Additional testing factors

Participants' information literacy and familiarity with academic library practices were a significant factor in the usability testing, although the extent to which it was so is unclear. For example, when looking at participants' scores for the known-item searching portion of the usability tests, it was difficult to know to what extent the ability to retrieve items was reflective of the tool's usability as opposed to the participant's online research skills. Similarly, it was unclear if the availability or location was apparent because the tool made it obvious, or because the user knew what to look for.

To retrieve locally held books, participants would need to know how to use call numbers; to order consortially held books, they would have to be aware of the existence of I-Share, the shared catalog and know how to use it to order materials. These operations are

inextricably linked to participants' level of research experience, if not their information literacy. It is also worth noting that the most deviation from the usability test script by the testers occurred when asking the retrieval questions. This suggests that this function of the tool was not easily understandable by participants. Deviation from the test script might also suggest that the language used in the usability metric itself was less than optimal. As an example, the prompt 1c (Appendix A) "Begin the process of retrieving this book as if you wanted to check it out" necessitated a clarifying paraphrase in a few instances so that participants would understand what they were being asked to do. Although each book-searching task referred to physical items on a shelf, the phrase "process of retrieving" seemed to cast some doubt in the participant's mind as to whether an electronic copy was somehow "retrievable."

There were also factors that contributed to the users' experience of the tools that were not related to the functions of the tools themselves. The least preferred of these was the link resolver, represented as a button labeled 'Find it @ UIC'. Participants were unclear about what it did and how it worked. Participants were also confounded by the intermediate page of the link resolver "360 Link," which is designed and hosted by Serials Solutions and has little resemblance to the discovery environment leading to it. While these concerns are not strictly speaking reflective of the usability of the tools under study, they no doubt contribute to the frustration some participants expressed at the research process in general.

LESSONS LEARNED

Lessons learned - study

Because the results of the study were used in part to make a time-sensitive decision about which discovery tool to retain, the test instrument was not as complete as it could have been, nor was there enough time to analyze the study data and make a decision about which tool to retain based solely on this data. In future studies, the objectives of the study should be uncoupled and more time given to each.

The usability test instrument also had some limitations. It emphasized known-item searching (four out of the six questions for each tool), despite the fact that discovery tools are used for many different types of searches. In addition to eliciting repetitive information, this emphasis also increased the length of the test. In future studies researchers might consider a different ratio of known-item searches to open-ended searches. Researchers could also create benchmarks for additional data points such as the questions asking the user to provide a number of the results they considered relevant. A multiple choice option from among several approximate percentages would generate more easily comparable data.

This study could also have benefited from a more consciously hybrid approach to the data analysis, i.e., creating functional sub-categories that tracked themes and trends emerging from the affinity matching exercise and allowing for more efficient analysis. In order to integrate more top-down analysis, more precise working hypotheses could help identify what researchers expect to see and thus enable them to balance question types according to these expectations.

Lessons learned – users

In addition to learning about their experiences with discovery systems, the researchers also gained a number of important insights with regard to our users. In some instances, the tasks researchers asked participants to complete were not necessarily tasks that participants

would be inclined to perform in a real-world scenario. This was particularly noticeable when researchers asked subjects to mark and e-mail a list of citations. A number of participants volunteered that they would not choose to do so and would instead choose other methods of access, from e-mailing entire articles to themselves to just reading the articles on the screen. As others have noted, testing often occurs in an artificial scenario that does not necessarily represent how end-users typically interact with these tools (Foster and MacDonald 2013). Among several options, a future study could be embedded in a credit-bearing library course or as part of the discipline-based curriculum, thereby approximating the “specific context of use,” a key element in the definition of usability (Barnum 2011, 279). Testing the usability of functions that users are unlikely to use is of limited utility outside of a testing scenario but nonetheless provides important guidance in deciding how to approach information literacy instruction involving these tools.

Study results support the contention that library-centered discovery tools, if not libraries themselves, are in stiff competition with Internet utilities such as Google and Amazon.com (Miller and Chad 2005; Majors 2012) for the hearts and minds of users. A number of study participants noted that they would use a Google product for certain searches rather than either of the tools, and a faculty member stated that she would go to Amazon.com to obtain a book, rather than pursue it through the options provided by the discovery tool. This suggests that while participants would use the tools under discussion in the future, the tools in many instances failed to meet a critical measure of usability: satisfaction. Unlike efficiency and effectiveness, the measure of satisfaction “is derived wholly from the user’s perception of satisfaction” (Barnum 20011, 12) and the participants’ preference for bypassing the tools in

favor of Internet utilities that perform the same functions indicates dissatisfaction with the tools.

The researchers observed problems with other online access routes to the library's resources that impaired the effectiveness of both discovery tools. For example, participants were often unclear about what happened when they clicked on the link resolver (Find it @ UIC) and how to obtain full-text items. As a result, even if the discovery tool led them to the link resolver correctly, they still did not find the full-text item.

The researchers also noted some general information literacy problems among users. Failure to complete tasks at times reflected a lack of understanding of the differences between material types or how to find an article within a journal. Interviewers adhered to the script while conducting the tests, but in some instances offered a brief instruction session after the test was completed to inform participants of valuable library services such as user-initiated consortial borrowing.

CONCLUSIONS

The working hypotheses of this study were that one of the discovery tools would prove more usable than the other and that one tool would be preferred over the other. The investigators also wanted to know the extent to which a correlation existed between the preferences of participants and the usability of the tools – as indicated by their “effectiveness”, “efficiency” and “satisfaction” for users – the critical measures of usability.

Analysis of the data revealed that WorldCat Local was the more usable of the tools, while study participants were evenly split in their preferences between Summon and

WorldCat Local. Participants also indicated, by a fairly wide margin, that if the library had to retain just one tool, they would want that tool to be Summon.

The decision to retain Summon as the Web-scale discovery tool was based on several factors in addition to the results of the usability study. One significant factor was the UIC Library's status as a customer of Serials Solutions Electronic Resource Management (ERM) tools, including 360 Link, the URL resolver. The interoperability of the ERM, 360 Link, and the discovery system allow a more seamless experience for users attempting to access electronic articles. This made it more likely that, although study results indicate that WorldCat Local was more usable, the library would nonetheless elect to retain Summon.

Because this study did not include an independent assessment of the information literacy of study participants, it is impossible to say with any degree of certainty how information literacy correlates with the usability findings. For example, the relative rankings of the groups in the study could reveal more than they currently do with a clearer idea of the participants' information literacy. Including an assessment of information literacy could be a future avenue for further research.

Although results are not statistically generalizable, the study sample was varied enough to allow researchers to extrapolate the kinds of issues end-users were likely to encounter while using these tools. The study findings therefore, can help to improve information literacy instruction by making clear what skills and issues need to be addressed in order for students to make optimal use of these discovery tools when conducting research. To the same end, the study results have already allowed for the customization of Summon to the extent currently possible.

ACKNOWLEDGMENTS

This study borrows from a study conducted by Melissa Becher and Kari Schmidt in 2011 for its coding scheme and organization. The authors would like to thank UIC librarians Lesley Brown, Sandra De Groote, Isabel Gonzalez-Smith, and Ryan Rafferty for their participation in the testing and other parts of the study.

REFERENCES

- Asher, Andrew D., Lynda M. Duke, and Suzanne Wilson. 2013. "Paths of Discovery: Comparing the Search Effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and Conventional Library Resources." *College & Research Libraries* 74 (5): 464-488.
- Barnum, Carol M. 2011. *Usability Testing Essentials: Ready, Set-- Test*. Burlington, MA: Morgan Kaufmann Publishers.
- Becher, Melissa, and Kari Schmidt. 2011. "Taking Discovery Systems for a Test Drive." *Journal of Web Librarianship* 5 (3): 199-219. DOI:10.1080/19322909.2011.583134.
- Bertot, John Carlo, Katy Berube, Peter Devereaux, Kerry Dhakal, Stephen Powers, and Jennie Ray. 2012. "Assessing the Usability of WorldCat Local: Findings and Considerations." *The Library* 82 (2). DOI: 10.1086/664588.
- Brantley, John S., Gwen M. Gregory, Jay Lambrecht, and Ling Wang. 2013. "Summon in the I-Share Environment." (Webinar). In CARLI Discovery Layer Webinar Series. Retrieved from <http://www.carli.illinois.edu/opac-team-webinar-summon-i-share>.
- Chad, Ken, and Paul Miller. 2005. "Do Libraries Matter." *The Rise of Library 2.0*. (White Paper). Retrieved from http://www.capitalibraries.co.uk/downloads/white_papers/DoLibrariesMatter.pdf
- Fahey, Sue, Shannon Gordon, and Crystal Rose. 2011. "Seeing Double at Memorial University: Two WorldCat Local Usability Studies." *Partnership: The Canadian Journal of Library & Information Practice & Research* 6 (2).

- Foster, Anita K., and Jean B. MacDonald. 2013. "A Tale of Two Discoveries: Comparing the Usability of Summon and EBSCO Discovery Service." *Journal of Web Librarianship* 7 (1): 1-19. DOI: 10.1080/19322909.2013.757936.
- Glaser, Barney G., and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.
- Gorman, G.E., and Peter Clayton. 2005. *Qualitative Research for the Information Professional: A Practical Handbook*. London: Facet Publishing.
- Gross, Julia, and Lutie Sheridan. 2011. "Web Scale Discovery: The User Experience." *New Library World* 112 (5/6): 236-247. DOI: 10.1108/03074801111136275
- Majors, Rice. 2012. "Comparative User Experiences of Next-Generation Catalogue Interfaces." *Library Trends* 61 (1): 186-207. DOI: 10.1353/lib.2012.0029.
- Nagy, Andrew. 2014. "Summon Product Update." (Webinar). In *Serials Solutions Product Update Series*. Retrieved from <https://serialssolutions.Webex.com/serialssolutions/ldr.php?RCID=2f32bd6c2a4f02f51a55f8e8bb60045>
- Vaughan, Jason. 2011. *Web Scale Discovery Services: A Library Technology Report*. Chicago, IL: American Library Association.
- Way, Doug. 2010. "The Impact of Web-Scale Discovery on the Use of a Library Collection." *Serials Review* 36 (4): 214-220. DOI: 10.1016/j.serrev.2010.07.002

APPENDIX A

Instrument: Pre-task questions, Interviewer-Administered Questions to Participants, Post-task questions

PRE-TEST QUESTIONS

1. What online tools (up to three, more if you like) do you use most frequently to find research articles and books? (for example Google, Google Scholar, library catalog, JSTOR, PsycInfo).

How often do you use these tools?

2. Have you used Summon or WorldCat Local before?

____ Summon

____ WorldCat Local

____ I have not used either

____ Don't know

3. (ASK ONLY FOR UNDERGRADS) Are you a Freshman ____ Sophomore ____ Junior ____ Senior ____

(if you are somewhere in between these levels, choose the one to which you are the closest).

4. What is your major/department?

(ASK ONLY FOR UNDERGRADS) What is your minor (if you have one)?

5. In your time at UIC, have you ever taken any library research classes or workshops either on your own or as part of another class?

Yes ____ No ____

6. How do you rate your Internet searching skills?

Excellent __ Good __ Fair __ Novice __

USABILITY TEST QUESTIONS - SUMMON

1. Using the Summon search tab on the library home page, search for the book The corporation: the pathological pursuit of profit and power by Joel Bakan.

1a. Is it available at UIC?

1b. If not, where is it available?

1c. Begin the process of retrieving this book as if you wanted to check it out.

2. Using the search box at the top of the Summon page, perform a new search for the book Pride and Prejudice and Zombies by Steve Hockensmith and Patrick Arrasmith.

2a. Is it available at UIC?

2b. If not, where is it available?

2c. Begin the process of retrieving this book as if you wanted to check it out.

3. Perform a new search for the book Beneath the Empire of the Birds by Carl Watson.

3a. Is it available at UIC?

3b. If not, where is it available?

3c. Begin the process of retrieving this book as if you wanted to check it out.

4. Task scenario: You need to find the following article to read for a class:

Daryl D. Green, and Gary E. Roberts. "Transformational Leadership in a Postmodern World: The Presidential Election of Barack Obama." *Academy of Strategic Management Journal* 11.1 (2012).

4a. Can you get the whole article online?

4b. If it is not available online, how can you get the entire article?

5. Search for information on sustainable energy consumption by municipal governments.

5a. How many results seem relevant to the topic? [NO BENCHMARK]

5b. Limit the results so that they are more relevant to your understanding of the topic.

5c. Limit your results to items published since 2000.

5d. From those limited results, choose three articles to email to yourself.

6. You are assigned to write a research paper on a topic of interest to you. Search for materials on this topic in Summon using whatever terms you want to and whatever features you choose.

Select some of these from the first page that you might use in a paper.

6a. What kind of publications have you chosen? (i.e., are these books, journal articles, conference proceedings, newspaper articles, broadcast transcripts?)

6b. Why did you select these materials? [NO BENCHMARK]

REFLECTION QUESTIONS

Now that you have performed several searches in Summon, reflect on the experience.

Which aspects of Summon did you find to be the most useful?

Which aspects of Summon did you find to be frustrating or confusing?

Navigate back to the UIC Library's homepage.

USABILITY TEST QUESTIONS – WORLDCAT LOCAL

1. Using the WorldCat Local search tab on the library home page, search for the book Why do ruling classes fear history?, and other questions by Harvey Kaye.

1a. Is it available at UIC?

1b. If not, where is it available?

1c. Begin the process of retrieving this book as if you wanted to check it out.

2. Using the search box at the top of the WorldCat Local page, perform a new search for the book Sense and sensibility and sea monsters by Ben Winters.

2a. Is it available at UIC

2b. If not, where is it available?

2c. Begin the process of retrieving this book as if you wanted to check it out.

3. Perform a new search for the book Painters of the Ashcan School: the Immortal Eight by Bennard Perlman.

3a. Is it available at UIC

3b. If not, where is it available?

3c. Begin the process of retrieving this book as if you wanted to check it out.

4. Task scenario: You need to find the following article to read for a class: Lorenzo-Dus, N. "Real Disorder in the Court: an Investigation of Conflict Talk in Us Television Courtroom Shows." *Media, Culture and Society*. 30.1 (2008): 81-107.

4a. Can you get the whole article online?

4b. If it is not available online, how can you get the entire article?

5. Search for information on the credibility of child eyewitness testimony.

5a. How many results seem relevant to the topic? [NO BENCHMARK]

5b. Limit the results so that they are more relevant to your understanding of the topic.

5c. Limit your results to items published since 2000.

5d. From those limited results, choose three articles to email to yourself.

6. You are assigned to write a research paper on a topic of interest to you. Search for materials on this topic in WorldCat Local using whatever terms you want to and whatever features you choose. Select some of these from the first page that you might use in a paper.

6a. What kind of publications have you chosen? (i.e., are these books, journal articles, conference proceedings, newspaper articles, broadcast transcripts?)

6b. Why did you select these materials? [NO BENCHMARK]

REFLECTION QUESTIONS

Now that you have performed several searches in WorldCat Local, reflect on the experience.

Which aspects of WorldCat Local did you find to be the most useful?

Which aspects of WorldCat Local did you find to be frustrating or confusing?

POST-TEST QUESTIONS

1. In your own words, what is the purpose of these tools?
2. Which tool did you prefer? Why?
3. Which tool provided the best results?
Why did you like these results better?
4. Which tool was easier to use? Why?
5. Would you use either of these tools again? Why or why not?
6. If the library could only keep one of these tools, which would you choose?