Use and Usability of a Discovery Tool in an Academic Library

Authors

Scott Hanrath
Miloche Kottman

This is an Accepted Manuscript of an article published by Taylor & Francis in The Journal of Web Librarianship *on 2015-02-23, available online:*
http://dx.doi.org/10.1080/19322909.2014.983259

Abstract

In order to assess the use and usability of a new discovery tool, staff at the University of Kansas Libraries conducted usability tests with twenty-seven users and analyzed three semesters of the tool's usage as measured by custom event tracking implemented in Google Analytics and usage statistics drawn from the discovery tool and server logs. An initial study with sixteen users was conducted prior to launching the new tool, and a subsequent study with eleven users was conducted a semester after the launch. This article describes test participants' success using the new tool to complete basic library research tasks, details the specific features they used in their attempts (e.g., facets, "did you mean" suggestions), and identifies areas where changes were made to address problems identified in the studies, including changes outside the tool itself. In addition, comparisons between feature use in the discovery system as observed in usability testing and feature use as measured by event tracking and log analysis are discussed, including implications for the design of future tests.

Acknowledgements

The authors wish to thank Judith Emde for her part in planning and conducting the user studies described in this paper.

Keywords

Usability, Web analytics, Web-scale discovery, Academic libraries

Introduction

In 2012, the University of Kansas Libraries (KU Libraries) purchased and implemented the discovery interface Primo by Ex Libris along with Primo Central, a subscription index of online content. The KU Libraries' goal was to provide our users with a single search box and results that included journal articles as well as items from our catalog and digital collections. An implementation team was charged with launching the new discovery service and customizing the interface to meet the needs of our users.

Primo's user interface has several customizable features, for example the order of the facets that can be used to narrow the results set and the amount of bibliographic information presented on the results screen. To aid decision-making about customizations and to assist the reference and instruction staff of the KU Libraries in effectively supporting our users with this new tool, the implementation team pursued multiple methods of gathering information about how users approached our discovery interface.

The discovery interface was made available as a "soft launch" or preview during the summer of 2012 prior to the official fall semester launch. During the soft launch an online feedback form was available for users to submit comments. The implementation team also partnered with KU Libraries instructional staff to facilitate a series of workshops for library staff, which provided both structured and informal feedback.

Beyond these measures, which were either aimed specifically at KU Libraries staff or relied on users to approach us with comments, the authors conducted formal usability tests with

students. The tests were designed to achieve two goals. First, we sought to identify any major problems users might have using the discovery interface to complete common library research tasks. Second, we sought information about how users might approach new features that were not available in the previous online catalog or other search tools provided by KU Libraries.

An initial round of usability testing was conducted in the summer of 2012, during the soft launch period. A second round of tests was conducted during the Spring 2013 semester, in part because the implementation team made several interface changes and upgraded to a newer version of Primo after the initial Fall 2012 launch. The second round of tests allowed us to supplement what we had learned from the first tests, providing a clearer picture of how our users were approaching the discovery interface.

In addition to usability testing, the implementation team reviewed Web analytics to understand how the discovery interface was being used. Custom event tracking in Google Analytics was implemented to track the use of specific features, such as facets and "did you mean" suggestions, in addition to traditional analytics like page views and visits. Usage statistics provided by Primo itself as well as other server logs were also used to track feature use. The analysis of these statistics provided the authors with the opportunity to compare the behavior observed in our usability tests with the behavior of users outside of a formal test setting. These comparisons are discussed below as a rough method for gauging the representativeness of behavior observed in usability tests with a small number of users and as a potential source of input into changes for future usability test design.

Literature Review

Academic libraries have relatively recently turned to network-scale discovery systems to supplement or replace existing online catalogs and older federated search tools. These discovery

systems (which include OCLC's WorldCat Local, EBSCO's Discovery Service (EDS) and ProQuest's Summon) typically index content of a variety of types and from a variety of sources in an attempt to provide single-search box access to a large portion of a library's resources. As frequently noted at the outset of discussions of library discovery systems, Google has greatly impacted users' search expectations and provided much of the impetus behind libraries' adoption of these search tools (Lown, et al. 2013; Asher et al. 2012). Beyond simply consolidating resources in one place, Lorcan Dempsey (2012) notes that the single simple search boxes typical of discovery systems require a shift in approach from traditional library search behavior. Instead of being asked to "make choices up front by closely specifying a query," users are instead asked to "refine a result" through narrowing mechanisms such as facets or a "bento box" display in which results are grouped by type or source, or by following suggestions for additional searches (Dempsey 2012). Below we highlight several studies of discovery tools and discussions of Web analytics particularly relevant to our tests and methods.

Usability studies, a common practice in academic libraries since the advent of the World Wide Web, have included several tests of single discovery products and comparisons of user performance across multiple products. Xi Niu et al. (2014, 424) provide a useful summary of the major findings of a number of such studies and Aaron Nichols, et al. (2014, 175) provide summaries of three usability studies of Primo.

In a multiple product study, Andrew Asher, et al. (2012) compared student use of Summon, EDS, Google Scholar, and traditional library databases. They argue that the quality of the discovery tools provided to undergraduates, and especially the default relevancy rankings of the tools, warrants close examination by libraries. While undergraduate students have great confidence in their ability to surface quality results through simple keyword searches, there is

evidence that they are often ill-prepared to make the sophisticated evaluations of resource quality required by large, diverse results sets and that a library's selection of a particular discovery tool may greatly impact the resources its users select (Asher, et al. 2012, 473-474).

Tamar Sadeh (2008, 22-23) discusses an early usability study of Primo at the University of Minnesota where participants reported positive impressions of the discovery tool and of new features like facets, a theme common to subsequent studies. David Comeaux's (2012, 205) findings from user testing of Primo at Tulane University show that users rated the discovery system highly, even though they did experience some difficulty completing tasks using the system. Nichols, et al. likewise found that user attitudes towards Primo did not always match user behavior: in one case a user who used the tool effectively expressed confusion about the interface, while in other cases users described Primo positively yet they struggled to complete more complex tasks (2014, 187-188)

In discussing their usability study of EDS, Jody Condit Fagan, et al. (2012, 100-101) note that beyond uncovering specific usability problems, their results suggest that libraries should be mindful of how discovery systems integrate with other library resources and services, such as subject-specific databases and reference services.

Other studies of discovery systems have analyzed usage statistics and search query logs. Cory Lown, et al. (2013) analyzed two semesters of usage statistics from North Carolina State University's single-search box and bento box results, finding that focusing a single search box discovery tool too narrowly on articles and books may overlook user needs for other types of library resources or services. Further, they argue that on-going evaluation of discovery system use, particularly of the most frequently searched queries, is important to allow libraries to evolve their understanding of user needs (Lown et al, 2013). In their attempt to manually classify search

queries collected from the University of Michigan's single search box, Suzanne Chapman, et al. observed a high occurrence of known item searches (2013, 414). They suggest that the high occurrence of known item searches could be attributed to the fact that while the library is not a user's sole option for discovering resources, it may be a user's sole option for actually obtaining a resource once discovered (Chapman, et al. 2013, 415). Niu, et al. analyzed a month's worth of transaction logs to compare facet use and search query formation and reformation in Primo and VuFind. They found that facets appeared to be helpful to users based on frequent use and that searches for non-electronic resources tended to include more terms and more query submissions than searches for electronic resources (Niu, et al. 2014, 428-429).

Steven J. Turner (2010) and Wayne Loftus (2012) provide an overview of using Google Analytics to evaluate library Web sites. Though not focused on discovery systems, their examples are applicable to studies of discovery system use. Turner describes different reports available in Google Analytics and illustrates how concepts developed for e-commerce analytics, such as key performance indicators and conversion rates, can be adapted to assess library Web site effectiveness (2010). Loftus describes using custom event tracking to extend the types of data available in Google Analytics (2012, 10).

Methods

Usability Tests

We focused the usability testing of our Primo discovery interface on basic library searching scenarios for undergraduate and graduate students. Since Jakob Nielsen (2009) has demonstrated that a small number of users can find the most glaring problems, we planned to solicit between ten-fifteen volunteers for our usability tests. The first round of testing (Study 1), held in July 2012, included sixteen participants. This round of testing was performed prior to the

Fall 2012 launch of the new discovery interface. The second round of testing (Study 2), held in April 2013, included eleven participants, for a total of twenty-seven participants.

Participants were solicited via KU Libraries' social media account and by personally approaching students present in the library to request their participation. Recruiting students already present in the library proved more effective than recruiting students online in advance. Participants were offered university-branded water bottles as incentives during the first round and pre-paid university dining gift cards during the second round. (The dining gift cards generally elicited greater enthusiasm than the water bottles).

Participants were asked to complete five tasks in the first round of testing and four tasks in the second (see Appendix 1) with sessions anticipated to take no more than thirty minutes. Tasks for the second round of tests were altered slightly based on our experiences administering the first round tests, as detailed below. Beyond allowing us to observe any difficulties users may have had in completing basic library research tasks, the tasks presented opportunities to observe which features of the discovery interface were used, successfully or not, in our users' attempts to complete the tasks. For example, would users take advantage of the facets? Would they be able to understand the FRBR (Functional Requirements of Bibliographic Records)-like clustering of similar titles into a single item on the search results page?

Prior to beginning the tasks, participants answered six demographic questions. At the conclusion of the session participants completed a post-test survey that asked for their overall impression of the search tool (see Appendix 2).

TechSmith's Morae software was used to record and analyze the participants' responses. The Morae Recorder was used to capture the screen display, keystrokes, mouse clicks, audio and video of the participant. Following a "think-aloud" protocol, participants were encouraged to

verbalize their intentions as they completed tasks and were prompted to do so if they appeared stuck or were not providing verbal feedback. Morae Manager was used to code whether or not a participant successfully completed a task, calculate task completion time, capture search terms, and mark usage of various interface features by participant and task. Further analysis was conducted using Excel and SPSS.

Event Tracking and Usage Statistics

Custom JavaScript was added to our Primo front end to track user interactions with the search interface as Events in Google Analytics. Search queries, user clicks on facets, the presence of and clicks on "did you mean" suggestions, and other interactions were captured as events. Event data was analyzed using Google Analytics' reporting tools and exported for further analysis in Excel and SPSS. Event tracking data supplemented data on user behavior available via Primo's reporting system and server logs.

Demographics

The twenty-seven participants included twenty undergraduate students and seven graduate or professional students from a variety of majors. The number of times participants reported searching for library resources in the previous semester ranged from never to daily. The majority of participants in both studies (twenty-two of twenty-seven) indicated that they had not used Primo, which we referred to as the "New Library Search" on the questionnaire. However, Primo had been the default search on the KU Libraries homepage since August 2012, so it is possible that some Study 2 participants had used Primo without realizing it.

Results and Discussion

Known Item Searching

The first three tasks of the user testing protocol asked participants to conduct known item searches, where a title or other information was provided and participants were required to find the item in the discovery system, then provide additional information about it and/or access it directly. The prompts supplied to the participants were:

> Task 1. Do the KU Libraries hold a physical copy of the book, *James Joyce and the Politics of Egoism* by Jean-Michel Rabaté, published in 2001? Where would you locate it?
>
> Task 2. Please find an electronic version of the book, *The Scarlet Letter*, by Nathaniel Hawthorne. If available, please access the book." (The leading article "the" was omitted from the prompt for the first study).
>
> Task 3. Search for this scholarly journal article, 'On the tragedy of love in The Scarlet Letter'. Published in the journal *Studies in Literature and Language* in 2011. Pull up the full text of the article.

Task 1 was intended to test a user's ability to find a known physical item from our collection. We were further interested in the user's ability to use Primo's "View all versions" feature, which attempts to group together multiple versions of the same item into a single result (e.g., different editions or adaptations of the same work) in a manner similar to FRBR. We therefore selected an item which would present participants with a "View all version" link – in this case an item where a physical copy and an electronic copy were grouped into a single result.

Based on very early feedback from KU Libraries staff, the implementation team moved the "View all versions" link from Primo's default position on the far right of the screen to just underneath the title and brief bibliographic information.

Task 1 was administered identically in both rounds of testing. Table 1 summarizes the success rates for participants for Task 1 for both Study 1 and Study 2.

Table 1. Success Rates

Task	Number of successful completions (N=27)	Percent of Successful completions
Task 1	19	70.4%
Task 2	21	77.8%
Task 3	24	88.8%
Task 4	23	85.2%

More than half of the participants were successful in determining that KU Libraries did hold a physical copy of the item. Participants were rated as successful if they were able to perform a search, find the item in the results set, and provide the facilitator with the item's call number and building location.

Table 2 summarizes the time needed for participants to successfully complete Task 1. Participants were told at the beginning of each task to take as much time as they would like to complete the task. If a participant felt unable to complete the task using the discovery interface provided to them, they were to indicate to the facilitators that they were ready to move on to the next task.

Table 2. Successful Participant Time to Task Completion in Minutes

Task	Mean	Minimum	Maximum	Standard Deviation
Task 1 (N=19)	1.6	0.6	3.1	0.6
Task 2 (N=21)	2.2	0.7	6.3	1.5
Task 3 (N=8)	1.7	1.3	3.9	0.4

A majority of our participants were able to successfully complete the known item search in an average of just over 1.5 minutes. Unsuccessful participants gave up on the task after an average of 2.5 minutes.

The results from this task also suggest that the "View all versions" grouping of electronic and physical items did not ultimately impede users in finding a known item. However, some users did struggle to find information associated with the physical item in this grouping. In the version of Primo we used for the test (version 3.1.3), participants that clicked on the title of the grouped items in the brief result went directly to the online version of the resource. This has changed in subsequent releases of Primo, where clicking on the title now takes users to a "View all versions" page that separately lists brief results for all titles matching the search. Our results suggest that this was a beneficial change.

Several users, some who were ultimately successful in completing the task and others who were not, quickly clicked on the title that led to the online version. In fact, over half of all users in both studies clicked on the title at some point during the task (56 percent in Study 1, 64 percent in Study 2, and 59 percent overall). Even in only those cases where users were ultimately successful, almost half of users clicked on the title (47 percent). Further complicating matters, the ebrary page for the online version displayed an "LC Call Number" alongside other identifiers for the work. We observed users closely examining this section, clearly expecting to find information about KU Libraries' holdings, when in fact the only way to obtain that information was to return to the discovery interface. This is the first of several examples we observed where the design of systems beyond the discovery tool itself impacted users' ability to complete tasks.

Task 2 was intended to test a user's ability to find the online version of a known item in Primo. Nearly 80 percent of participants were able to successfully complete this task, a higher overall success rate than locating a physical item in Task 1 (see Table 1).

Successful participants on average spent more time on Task 2 compared to Task 1 and in some cases spent considerably longer, with the maximum time spent on Task 2 more than twice as long as the maximum for Task 1 (see Table 2). Interestingly, the average time spent by unsuccessful participants on Task 2 (1.7 minutes) was somewhat lower than the time spent on Task 1 (2.5 minutes).

Task 2 required the user to not only locate the electronic version of the item, but also to actually access the full text of the item. In Study 1, some of the extra time participants spent on this task was because of the poor relevancy ranking and out-of-the-box boosting of local resources in Primo. The prompt for Study 1 was to find the book, "Scarlet Letter" by Nathaniel Hawthorne. Consequently, most participants entered their search term as "scarlet letter." Unfortunately, the omission of the initial article "the" made a significant difference in the search results. The relevant result when searching "scarlet letter" was found near the bottom of the results page while the relevant result for the search "the scarlet letter" was the second result listed. Adding the author's name to the search terms, as several participants did, did not improve results, and in some cases made it worse. This is in keeping with Nichols, et. al. findings that participants who used more complex search queries in Primo were often less successful than those who used short, simple queries (2014, 184). When we repeated the study in Spring 2013, we added the initial article to the title prompt. We have since succeeded in boosting our local resources to appear higher in the results, and subsequent releases of Primo have attempted to

improve the relevance ranking regardless of the presence or absence of leading articles in the search string.

Participants also experienced difficulty in determining whether they had located the full text of the resource. This difficulty was not due to Primo but is worth noting. KU Libraries have Hawthorne's *The Scarlet Letter* available as full text on several different electronic platforms. While some platforms immediately display the full text, others open to a title screen and require an additional click to open the full text. Still other platforms provide a display that simply does not seem to meet users' expectations of a full-text book. For example, when an e-book page didn't immediately display the full text or a table of contents, some participants assumed they had opened an incorrect document.

Task 3 asked participants to conduct a known-item search for a scholarly research article, given the title, journal, and date of the article. Overall, among the three known item search tasks, participants experienced the greatest success in locating the journal article in Task 3, with only three out of twenty-seven participants failing to successfully complete the task (see Table 1).

In Study 1, as part of Task 3 participants were further asked to use two of Primo's more advanced features. They were asked to "Mark and save the citation to the e-shelf" (a feature of Primo that allows users to save citations in a list they may return to later) and to "email the citation to yourself." Participants experienced great difficultly completing this part of the task. In retrospect, combining these related, but distinct tasks into a single prompt made accurately measuring the time needed to complete the task impossible. For our second study we reduced the prompt for Task 3 to only include locating the item and accessing the full text. Because of this. in Table 2 we report the time to completion only for the participants in Study 2.

Not only were participants from both studies more successful in completing Task 3 compared to the other known item tasks, they spent less time on average completing the task. It is possible that the experience gained with Tasks 1 and 2 made it easier for participants to complete Task 3. In future studies randomizing the order of tasks may yield more conclusive results.

Though participants were more successful with this task, we did observe users struggling in two areas related to accessing the full-text article: using the link resolver page and navigating the resource platform. For most subscription resources, Primo provides an Open URL request link to our link resolver, not a direct link to an item. In many cases the full text was available on multiple platforms. In order to access the full text, users were required to select a platform from the list provided by the KU Libraries' Open URL Link Resolver, Serial Solutions' 360 Link. Several users successfully navigated from the discovery results list to the link resolver, but struggled to select a platform. Others successfully selected a platform from the link resolver, but had difficulty quickly finding the full text from amongst other information provided by the platform. We observed users struggling in particular with the first platform listed in our link resolver, where the path to the full text was obviously unclear to them. Several users were distracted by the prominent legend that defines the different icons used on the page and attempted to use the icons in the legend to retrieve the item. Other users returned to the link resolver, selected a different platform, and then quickly found the full-text item on the new platform.

For finding full-text resources as required in Tasks 2 and 3, aspects of the search process beyond the Primo discovery interface posed the greatest challenges to our participants. In order to access the full text, users were required to negotiate three different systems: the discovery

interface, the link resolver interface, and the interface of the source or platform for actually delivering the item. In our tests, the usability of the link resolver and the platforms accessed from there greatly impacted the success of the participants. In addition, the order in which platforms were listed on the link resolver page impacted users' ability to complete tasks, especially when the platforms listed first posed usability problems. Based on the results from Tasks 2 and 3, we both altered the design of our link resolver page to better highlight direct links to the article and adjusted the preferred order of the sources based on our observations of the usability of the sources.

The overall success rate for the known items search tasks 1, 2, and 3 was just under 80 percent. All of our known item search tasks required participants to not only find the item in the discovery system's results set, but to also either display the full text of the item or give the call number and building location of the item, requiring additional manipulation of the results and, in the case of Tasks 2 and 3, successful use of systems external to the local discovery interface. Evaluation of a user's ability to use a discovery system by itself may therefore be insufficient to gauge his/her ability to actually retrieve results.

Open Ended Searching

The remaining tasks of our first study required participants to find resources they deemed relevant to a topic. Task 4 supplied the topic and asked participants to "Search for an online, peer-reviewed article published in the last 5 years on the topic: U.S. foreign policy and immigration from Mexico." In Task 5, we asked participants to supply their own topic: "Tell us about an assignment in the last year where you needed sources. Use this search to find 2 resources that you think would have been appropriate." Participants were asked to make their own judgment about whether the resource they identified was relevant to the topic.

In Study 2, the supplied topic in Task 4 was changed to "drinking and academic achievement in college" and Task 5 was dropped from the tests. While it was beneficial to observe participants in this task, we found that participants' understanding of our open-ended task prompt and the types of assignments described by students varied so widely that it was difficult to compare their experiences in quantifiable terms of success and time. As such, Table 1 summarizes the success rates for only Task 4 from both studies.

The overall success rate for this open-ended search task was higher than the overall success rate for the known item search tasks (85.2 percent versus 77.8 percent) and in both studies exceeded 80 percent. Time to task completion was not measured for this task.

Feature Use

Facets

In addition to gauging our users' ability to successfully complete basic library research tasks, we were interested in which features of the discovery system users employed to complete their tasks. Table 3 summarize the use of Primo's facets by participants in our studies. Across all tasks, 35.2 percent of participants used a facet for a given task. That is, twenty-seven participants attempted four tasks each and a facet was used in twenty-six of the resulting 108 opportunities. The open-ended search in Task 4 yielded the greatest use of facets.

Table 3. Facet Use

	Number of Participants Who Used Facets	**Percent of Participants Who Used Facets**	**Total Number of Facet Clicks Across Participants**	**Facet Clicks Per Participant**
Task 1 (N=27)	4	14.8	5	0.2
Task 2 (N=27)	14	51.9	20	0.7
Task 3 (N=27)	5	18.5	7	0.3
All Known Item Tasks (N=71)	23	32.4	32	0.4
Task 4 (N=27)	15	55.6	50	1.9
All Tasks (N=108)	26	35.2	82	0.8

More than half (55.6 percent) of participants clicked on a facet in Task 4 compared with around one third (32.4 percent) of participants for the three known item tasks (32 facet usages out of 71 opportunities for 27 participants in three tasks). It is possible that the participants were simply more familiar with the available features after already attempting three tasks and therefore more likely to use the facets in subsequent tasks. Nichols, et.al. report a similar finding, that more users used facets for open-ended research tasks than other tasks and likewise presented users with an open-ended research task after they had already attempted several tasks (2014, 183-184). But facet use did fall from a relatively high 51.9 percent of participants using facets for Task 2 to 18.5 percent of participants using facets for Task 3, suggesting that facet use was not entirely dependent on previous experience with facets.

Further, as shown in Table 3, participants also made greater use of the facets in Task 4 compared to the other tasks. In Task 4 there were 1.9 facet clicks per participant compared to 0.4 clicks per participant for the known items tasks. In other words, more participants clicked on more facets in the open-ended Task 4 compared to any of the three known-item search tasks taken individually.

Because we have tracked facet clicks in Primo using Google Analytics since launching Primo, we are able to make a rough comparison between facet use in our formal usability tests and facet use by our general user population. Custom JavaScript was added to our Primo front end to track user clicks on facet links as Events in Google Analytics. For the first three semesters since our discovery interface became available, there were 68,467 unique facet click events. In Google Analytics the number of unique events is equivalent to a visit during which an Event occurred at least once (e.g., if a user clicks on a Facet twice during a visit, that yields only one unique facet click Event). During the same time period, there were 321,418 visits to our Primo site, yielding 0.2 unique facet click events per visit. If we use a Visit as a rough proxy for the tasks in our studies, we see that the number of facet clicks in our usage statistics (0.2) is closer to the facet use in the known item tasks of our study (0.4) than to the open-ended search tasks (1.9). The comparison is not perfect (e.g., users may, of course, complete more than one task in a session and a greater variety of users utilize our discovery system than were represented in our tests). But the comparison is useful in that it suggests that the behavior we observed in our formal usability tests was not wildly disimilar from what we subsequently observed in routine use of our discovery system.

Search types

The number and type of searches performed by participants were also tracked, along with the length of the search queries. Primo includes both a default keyword "basic" search as well as an "advanced" search option that includes a number of different limiting options (e.g., search only in the title, author, or subject field) that may be combined. As we implemented Primo, we had a number of potential decisions to make about how to configure, display, and market the

advanced search options. Knowing how likely users were to avail themselves of the advanced search would help us determine how best to focus our efforts.

Table 4 summarizes how often participants in our studies used the default basic search and the advanced search. The open-ended Task 4 yielded the highest number of average total searches, whether basic or advanced, at 2.6 and the highest average number of advanced searches at 0.9. Task 2, the known item article search, yielded the highest number of basic searches at 1.8. Overall, participants in our tasks averaged not quite 2 searches per task at 1.8.

Most of the searches were basic searches. For the known item searches, Tasks 1-3, advanced searches comprised 9 percent of total searches. For Task 4, advanced searches comprised 35 percent of searches. Overall, advanced searches comprised 18 percent of total searches.

Table 4. Number of Searches Performed

	Basic		**Advanced**		**Total**	
	Number	**Mean**	**Number**	**Mean**	**Number**	**Mean**
Task 1 (N=27)	31	1.1	3	0.1	34	1.2
Task 2 (N=27)	48	1.8	3	0.1	51	1.9
Task 3 (N=27)	36	1.3	5	0.2	41	1.5
All Known Item Tasks (N=71)	115	1.6	11	0.2	126	1.5

The average number of advanced searches performed for the known item tasks is comparable to the number of advanced searches performed in Primo as measured by its usage statistics. For the first three semesters Primo was available to our users, Primo's statistics show a total of 826,713 sessions (Primo and Google Analytics measure sessions differently). A Primo session represents one or more requests from a single user without a long gap in between. During the same time period Primo logged 154,617 advanced searches, or an average of 0.2 searches per session. If we use Primo sessions as a proxy for user tasks, as we did with Google Analytics

visits when comparing facet use, we see that our usage statics for advanced searches matches the 0.2 advanced searches averaged for Tasks 1 through 3 and is considerably lower than the 0.9 advanced searches averaged for Task 4. As with facet usage, the average number of advanced searches logged in our usage statistics is closer to the average number of advanced searches in the known item tasks than the open-ended search task during our usability tests. Given the relatively low use of advanced searches observed in our tests, our implementation team concluded that changes to the advanced search experience would have relatively low impact and chose to focus our initial efforts on improving the basic search experience. Analysis of usage data confirmed that decision.

Table 5 summarizes the average length of search queries in all tasks compared to the average length of queries as measured in usage statistics. While there is considerable variation within the three known item tasks in both the number of search terms used and the number of total characters in the query, the overall averages for the known item tasks (6.3 terms and 38.7 characters) are quite similar to the averages for the open-ended Task 4 (5.6 terms and 38.3 characters) and the averages for all tasks (6.2 terms and 38.4 characters). Analysis of a combination of log files (for Primo searches that originated from the Libraries Web site homepage via a redirect script) and Events tracked with Google Analytics (for subsequent searches within Primo) shows that for the first three semesters Primo was available to our users, the average number of search terms in a query was 4.7, while the average length in characters was 33.1. These averages are lower than observed in testing, but suggest that the behavior observed in our tests was not radically dissimilar from routine use. In contrast to facet use and number of searches performed, the length of search queries observed during actual use was closer to our open-ended search task than our known item search tasks.

Table 5. Average length of search queries

	Mean number of terms	**Mean number of characters**
Task 1 (N=34)	6.7	37.5
Task 2 (N=51)	3.8	28.1
Task 3 (N=41)	10	52.3
All Known Item Tasks (N=124)	6.3	38.7
Task 4 (N=71)	5.6	38.3
All Tasks (n=197)	6.2	38.4

Boolean and phrase searching.

In part because KU Libraries staff raised concerns in staff workshops about Primo, the percent of searches using Boolean operators and phrase searching (where quotes are used to group terms in a query) were tracked in our tests. Primo supports Boolean searching through the use of Boolean operators in all capital letters (e.g., "AND"). Only queries where the operator was capitalized were counted as Boolean searches. Table 6 summarizes the results.

Table 6. Percent of queries using Boolean Operators or Phrase Searching

	Percent of Queries using Boolean Operators	**Percent of Queries Using Phrases**
Task 1 (N=34)	5.9	2.9
Task 2 (N=51)	3.9	3.9
Task 3 (N=41)	0	34.1
All Known Item (N=124)	3.2	13.7
Task 4 (N=71)	2.8	4.2
Total (N=197)	3	10.2

The known item search tasks yielded a slightly greater precentage of Boolean searches than the open-ended search task (3.2 percent compared to 2.8 percent). Overall, use of Boolean searches in our tests was low at only 3 percent of searches. Analyzing search logs and Google Analytics events from our discovery systems after it launched also shows that Boolean searches

were seldom performed, at only 3 percent of searches, and suggests that the Boolean search behavior we observed in our tests was typical. Given the low usage of Boolean searches, we opted not to emphasize Boolean searching in additional help documentation or prompts, as several workshop participants had suggested. Instead, we developed an experimental feature that attempts to prompt users to use capitalized Boolean operators only upon submitting a query that the feature evaluates as likely to be a Boolean search (e.g., queries that contained lowercase Boolean operators that do not appear to be titles).

Note that Task 3 yielded no Boolean searches at all. However, it yielded a far greater percentage of phrase searches than the other tasks. This is perhaps because the item called for in Task 3 had a longer title than the titles in Task 1 and 2 or, it may be because the article title was presented to participants within quotation marks. Users may have simply transferred the quotation marks into their queries.

Overall in our tests, phrase searching was used in just over 10 percent of queries. If Task 3 is excluded, the percent of queries using phrase searching drops to just 3.8 percent, which is much closer to the percentage for Tasks 1, 2, and 4 individually and to the 2.7 percent of queries using phrase searching in the first three semesters of use of our discovery tool. In this case there is evidence that our choice and presentation of tasks yielded behavior considerabley different from that observed in routine use.

"Did you mean" suggestions

As is common in commercial search engines and library discovery products, Primo includes a "did you mean" (DYM) feature. The DYM feature offers suggestions for new searches based on possible misspelled terms in the original search query or in cases where the original search query generated few results but a similar query would yield more results. As this

feature was not offered by our existing catalog or federated search interfaces we tracked how often DYM suggestions were presented to participants during our tests and how often users clicked on the links for the suggested new search.

Table 7 summarizes the DYM suggestions generated and clicked on by task. Use of the DYM feature was low, totaling only eight clicks. The DYM suggestions were not nearly as likely to be used as the facets, which received eighty-two clicks. DYM suggestions were offered for 18 percent of the searches for known item tasks and users clicked on fewer than 10 percent of those suggestions. The open-ended search task saw a higher percentage of searches with DYM suggestions (33 percent) and a much higher percentage of those suggestions were clicked on (40 percent). In two semesters of actual use -- data are not available for the first semester -- users clicked on 8.9 percent of the DYM suggestions generated by Primo. As with use of facets and advanced searches, this rate of use of DYM suggestions is more similar to that observed for our known item tasks than for our open-ended task. Overall in our tests, DYM suggestions were clicked on at a rate of 0.07 clicks per task attempted. In two semesters of tracked usage, DYM suggestions were clicked at a rate of 0.1 clicks per visit. If we again use visits to the discovery service as a proxy for tasks in a testing situation we find the rates of DYM usage similar enough to have some confidence that the use of the DYM feature in our tests did not vary greatly from normal user behavior. It is important to note that Primo's DYM settings are configurable and that for our tests and most of the two semesters of data reported here we used the default settings. Changes to those settings may impact use and would require additional study.

Table 7. "Did you mean" (DYM) feature use

	Number of DYM suggestions offered	**Percent of searches with DYM suggestions**	**Number of DYM suggestions clicked**	**Percent of DYM suggestions clicked on**
Task 1 (N=34)	3	10%	0	0%
Task 2 (N=51)	13	27%	1	7.7%
Task 3 (N=41)	5	14%	1	20%
All Known Item Tasks (N=124)	21	18%	2	9.5%
Task 4 (N=71)	15	33%	6	40%
All Tasks (N=197)	36	22%	8	22.2%

Post-summary

When participants were asked about their overall impression of the new search tool, the majority of responses were positive in both studies (over 80 percent rated their impression as "Positive" or "Very Positive"). When asked how likely they were to use the search in the future, the studies differ slightly. More participants in the second study were "Very Likely" to use Primo in the future (55 percent) than were participants in the first study (31 percent).

During the second study participants were asked if they were likely to use the "Like" button to share results on Facebook. All eleven participants said "no" (with some commenting that they no longer use Facebook).

Participants in the second study were also asked if they were likely to conduct library research on a mobile device. Two participants said "yes" and nine said "no".

Conclusions and Future Research

The success rates for tasks in our two usability studies of the Primo library discovery system ranged from 70 percent to 88 percent, with an overall success rate of 80 percent across all tasks. Because the tasks were selected to represent common library tasks, these numbers are

perhaps lower than hoped for. When evaluating the overall task success rate, two factors should be considered: aspects of discovery outside the library search interface and skill level of the users.

First, as noted, several of our tasks required users to obtain the full online text of a resource. In the case of a full-text article, this required the participant to effectively use the discovery system, a link resolver, and a publisher or aggregator's Web interface. We observed participants experiencing difficulty with each of these components. And while other discovery systems may bypass the link resolver and link directly to the first available source for an article, this does not mitigate potential problems with publisher Web sites. It is therefore difficult to separate issues of delivery from issues of discovery. As noted in OCLC's study of user and librarian expectations of online catalog interfaces, "for many end users, without the delivery of something he or she wants or needs, discovery alone is a waste of time" (OCLC 2009, 20). If our tasks had focused more narrowly on just the discovery system and not its integration with delivery systems, the success rates would have almost certainly been higher. But that would have required creating distinctions between systems that our users, who likely view all of these components as simply part of "the library," might not recognize. Further, if we had adopted that narrower focus we would not have uncovered some specific usability issues experienced by our users. Matters of delivery impact the discovery system's value and are therefore important components to evaluating the system's overall utility. And though we did not enter our study of Primo with the intention of making changes to our external link resolver, we found it necessary to do so to address problems we observed. Future studies of discovery systems usability and usage could benefit from a more explicit focus on how discovery systems are integrated with other library services.

Second, there was considerable variation in the success rates of participants from our two studies, conducted several months apart. Success rates for the participants in Study 1 ranged from 81 percent to 94 percent, while for Study 2, the range was 55 percent to 82 percent. Despite what might be gauged as low success rates, more than 80 percent of participants across both studies responded that they felt "Positive" or "Very Positive" about the discovery system after completing the test. Asher, et al. argue that there is a continued need for ongoing training for students using library search systems. They remark, "well-prepared students can effectively use a variety of search tools, while poorly-prepared students will likely struggle with even the best-designed tools" (Asher, et al. 2012, 476). Our experience was very much in line with this assessment. In our pre-test questionnaire, more than half (55 percent) of participants indicated they were "Confident" or "Very Confident" in their ability to do library research. The inclusion of ratings of student preparedness based on methods other than self-assessment and their impact on student's ability to successfully use discovery systems could enrich future studies.

The analysis of usage data, particularly of how often new or updated features such as facets or "did you mean" suggestions were used, proved valuable for assessing whether or not the behavior we observed in user testing was "performative" in the artificial context of our test. Using a visit or session as proxy for a usability study task is a useful, though imperfect method of making this assessment. Possibilities for expanding and improving on this method include limiting the usage statistics to visits or sessions in which pre-defined successful actions occurred, for example limiting to sessions in which a link to a full text article was clicked on or a hold request for a book was made. Such an approach could make use of techniques described by Turner (2010) for defining, configuring, and measuring key performance indicators in Google Analytics to measure goals predefined by a library.

Continuing analysis of usage data also provides a means of monitoring how changes to discovery systems impact user behavior over time. As more software providers move to Agile Development cycles, a software development method that promotes more frequent releases of new versions and upgraded features, the need to monitor usage statistics becomes continuous. Establishing usage metrics to be monitored over time based on behavior observed in usability testing, or on changes made to the discovery system intended to ameliorate problems observed, could help inform decisions about when a new round of usability tests is needed to provide context not available from search logs or event tracking. The investments made by libraries in discovery systems call for a sustained program of assessment in which multiple modes of evaluation, including user testing and usage analysis, are applied in concert to improve the user experience.

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): 464-88.

Chapman, Suzanne, Shevon Desai, Kat Hagedorn, Ken Varnum, Sonali Mishra, and Julie Piacentine. 2013. "Manually Classifying User Search Queries on an Academic Library Web Site." *Journal of Web Librarianship* 7 (4): 401–21. doi:10.1080/19322909.2013.842096.

Comeaux, David. 2012. "Usability Testing of a Web-Scale Discovery System at an Academic Library." *College & Undergraduate Libraries* 19 (2): 189–206. doi:10.1080/10691316.2012.695671.

Dempsey, Lorcan. 2012. "Thirteen Ways of Looking at Libraries, Discovery, and the Catalog: Scale, Workflow, Attention." *EDUCAUSE Review*. http://www.educause.edu/ero/article/thirteen-ways-looking-libraries-discovery-and-catalog-scale-workflow-attention.

Fagan, Jody Condit, Meris A. Mandernach, Carl S. Nelson, Jonathan R. Paulo, and Grover Saunders. 2012. "Usability Test Results for a Discovery Tool in an Academic Library." *Information Technology and Libraries* 31 (1): 83–112. doi:10.6017/ital.v31i1.1855.

Loftus, Wayne. 2012. "Demonstrating Success: Web Analytics and Continuous Improvement." *Journal of Web Librarianship* 6 (1): 45–55. doi:10.1080/19322909.2012.651416.

Lown, Cory, Tito Sierra, and Josh Boyer. 2013. "How Users Search the Library from a Single Search Box." *College & Research Libraries* 74 (3): 227–41.

OCLC. 2009. "Online Catalogs: What Users and Librarians Want." https://oclc.org/content/dam/oclc/reports/onlinecatalogs/fullreport.pdf

Nichols, Aaron, Amber Billey, Peter Spitzform, Alice Stokes, and Catherine Tran. 2014. "Kicking the Tires: A Usability Study of the Primo Discovery Tool." *Journal of Web Librarianship* 8 (2): 172–95. doi:10.1080/19322909.2014.903133.

Nielsen, Jakob. 2009. "Why You Only Need to Test with 5 Users." http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/.

Niu, Xi, Tao Zhang, and Hsin-liang Chen. 2014. "Study of User Search Activities With Two Discovery Tools at an Academic Library." *International Journal of Human-Computer Interaction* 30 (5): 422–33. doi:10.1080/10447318.2013.873281.

Sadeh, Tamar. 2008. "User Experience in the Library: a Case Study." *New Library World* 109 (1/2): 7–24. doi:10.1108/03074800810845976.

Turner, Steven J. 2010. "Website Statistics 2.0: Using Google Analytics to Measure Library Website Effectiveness." *Technical Services Quarterly* 27 (3): 261–78. doi:10.1080/07317131003765910.

Appendices

Appendix 1 – User tasks

Study 1

Task #1: Do the KU Libraries hold a physical copy of the book, James Joyce and the politics of egoism by Jean-Michel Rabaté, published in 2001? Where would you locate it?

Task #2: Please find an electronic version of the book, Scarlet Letter, by Nathaniel Hawthorne. If available, please access the book.

Task #3: Search for this scholarly journal article, "On the tragedy of love in The Scarlet Letter". Published in the journal, Studies in Literature and Language. 2011. Mark and save the citation to the e-shelf, email the citation to yourself and pull up the article.

Task #4: Search for an online, peer-reviewed article published in the last 5 years on the topic: U.S. foreign policy and immigration from Mexico.

Task #5: Tell us about an assignment in the last year where you needed sources. Use this search to find 2 resources that you think would have been appropriate.

Study 2

Task #1: Do the KU Libraries hold a physical copy of the book, James Joyce and the politics of egoism by Jean-Michel Rabaté, published in 2001? Where would you locate it?

Task #2: Please find an electronic version of the book, The Scarlet Letter, by Nathaniel Hawthorne. If available, please access the book.

Task #3: Search for this scholarly journal article, "On the tragedy of love in The Scarlet Letter". Published in the journal, Studies in Literature and Language. 2011. Pull up the full-text of the article.

Task #4: Search for an online, peer-reviewed article published in the last 5 years on the topic: drinking and academic achievement in college. Why did you select this article? Please email the citation to [librarian email].

Appendix 2

PRE-TEST

Question #1: Student status – Undergraduate or Graduate. If undergraduate, what year are you in?

Question #2: What is your major?

Question #3: How often did you search for library resources last semester? Never, Once or twice a year, Once or twice a semester, Monthly, Weekly, Daily

Question #4: How confident are you in your ability to do library research? Not at all confident, A little confident, Somewhat confident, Confident, Very confident.

Question #5: Have you tried the New Library Search yet? Yes No.

POST-TEST

Question #1: What is your overall impression? Not at all positive, A little positive, Somewhat positive, Positive, Very positive.

Question #2: Did you find anything unclear or confusing?

Question #3: What one thing would you change?

Question #4: How likely are you to use this search in the future? Not at all likely, A little likely, Somewhat likely, Likely, Very likely.

Study 2 additional questions:

Question #5: Are you likely to use the "Like" button to share results on Facebook?

Question #6: Are you likely to conduct library research on a mobile device?