Paths of Discovery: Comparing the Search Effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and Conventional Library Resources

Andrew D. Asher, Lynda M. Duke, and Suzanne Wilson

In 2011, researchers at Bucknell University and Illinois Wesleyan University compared the search efficacy of Serial Solutions Summon, EBSCO Discovery Service, Google Scholar, and conventional library databases. Using a mixed-methods approach, qualitative and quantitative data were gathered on students' usage of these tools. Regardless of the search system, students exhibited a marked inability to effectively evaluate sources and a heavy reliance on default search settings. This article describes these results and makes recommendations for libraries considering these tools.

It would be difficult to overstate the impact that Google has had on searchers' experiences and expectations in the last decade. Google's ramifications are discussed relentlessly in the world of libraries and education and have been documented in myriad places.[1] Within the library, faculty and students have come to expect a simplified, fast, all-inclusive, and principally online research experience that mirrors their use of Google and other search engines. Increasingly, library faculty and staff have stressed the need to have "a single point of entry" or a "Google-like interface" for library databases if there is to be any hope of students and researchers consistently accessing library resources and maintaining the relevance of libraries in academia.[2]

Discovery tools are the latest attempt to address this need. These tools make it possible to create a centralized index of an institution's information resources and are designed so that a single point of access leads to a wide range of library content through a Google-style search box. Discovery tools have garnered a great deal of attention as libraries continually strive to streamline their online search functions, and competing examples of these tools have been implemented by a

Andrew D. Asher is Assessment Librarian in the Indiana University Libraries at Indiana University Bloomington; e-mail: asherand@indiana.edu. Lynda M. Duke is Academic Outreach Librarian, Associate Professor, and Suzanne Wilson is Library Technology and Resources Director, in The Ames Library at Illinois Wesleyan University; e-mail: lduke@iwu.edu, swilson@iwu.edu. © 2013 Andrew D. Asher, Lynda M. Duke, and Suzanne Wilson, Attribution-NonCommercial (http://creativecommons.org/licenses/by-nc/3.0/) CC BY-NC

growing number of academic libraries, including the libraries at the two universities involved in this study: Illinois Wesleyan University (IWU), which trialed the EBSCO Discovery Service (EDS) in the spring of 2011 and signed a three-year contract in the fall of 2011; and Bucknell University, which contracted with Serials Solution in November 2009 for a three-year commitment to use the Summon discovery service.[3] In spring of 2011, researchers at Bucknell and IWU joined forces to compare the efficacy of EDS, Summon, Google Scholar, and "conventional" library search tools on research tasks typically faced by undergraduate students. The purpose of this study was not only to test how these different search tools perform and function for students, but also to obtain a more holistic and user-centered understanding of student research practices to identify and address unmet student needs and instructional requirements.

Literature Review

Federated search tools appeared in the early 2000s, as an initial attempt to compete with Google.[4] Designed to simultaneously query multiple research databases from a single entry point, federated search tools' limitations are well documented, including long waiting times (particularly compared to Google), an inability to refine searches to the desired degree, problematic interfaces, and results lists that are difficult to use and interpret.[5] Relevancy ranking is also problematic when running parallel searches in multiple databases; more recently, federated searching has come under attack for not being compatible with smartphones or other mobile technology.[6]

Google Scholar (GS), launched in 2004, upped the ante. GS describes itself as providing "…a simple way to broadly search for scholarly literature. From one place, you can search across many disciplines and sources: articles, theses, books, abstracts and court opinions, from academic publishers, professional societies, online repositories, universities and other web sites."[7] Increasingly, libraries have chosen to make their link resolver and associated full-text available through GS. This practice's appeal lies in linking to the familiar "Google" pages with their simple search interface, thereby increasing the potential to draw users to library databases. However, criticisms of this tool include limited advanced search functionality, incomplete or inaccurate metadata, inflated citation counts, lack of usage statistics, and inconsistent coverage across disciplines.[8] There is also a lack of clarity regarding what GS actually indexes, and students are often unaware that GS's preferences must be manually set to link to libraries' resources. Because of this, students are often confused by the request to pay for articles or the need to click through to a library website.[9]

Discovery tools came to fruition in 2007 with OCLC's WorldCat Local, followed in mid-2009 by Serials Solutions Summon. In 2010, EBSCO Discovery Service (EDS), Innovative Interfaces Encore Synergy, and Ex Libris Primo Central entered the arena.[10] Competition remains intense among these vendors, and there is ongoing discussion and debate about the strengths and weaknesses of each product.[11] Providing "seamless" searching capabilities across a variety of databases, these tools have been heralded as the library's answer to Google. By preharvesting content from myriad databases into a single index, these tools improve on federated searching tools' speed, deduplication abilities, relevancy rankings, and the amount of content that can be accessed.[12]

Given the relatively recent development of discovery tools, little has been written about how the tools actually perform for users, and, at the time of writing, the authors were unable to identify any published user experience study comparing GS, Summon, and EDS discovery tools with each other or with library databases in general.[13] However, some usability studies of individual discovery tools have been completed. Julia

Gross and Lutie Sheridan, in a small usability study of Summon, determined that students were overwhelmingly drawn to using the single search box and it became the preferred navigation path.[14] In the fall of 2010, Sarah Williams and Anita Foster conducted six usability sessions with EDS. All participants felt it was a "useful tool for actual research or coursework" and they would be likely to use it again.[15] David Howard and Constance Weibrands explore the librarian's response to implementing a web scale discovery tool and the philosophical shift necessary to embrace the new technology, including facing fears of "dumbing down" the search process, rethinking educational materials, and trusting that the inner workings of Summon are reliable.[16]

Various articles have been written providing information on implementation, product overviews, structure, pricing, ability to customize, types and amount of content available, and availability of usage statistics.[17] This study seeks to move beyond technical issues and single-tool evaluations to make a more comprehensive investigation that compares how students use different search tools and the types of materials they discover during their searches.

Research Design
This study used a mixed-methods approach to gather both quantitative and qualitative data on students' usage of search tools. Participants in the study were assigned to one of five test groups based on the search tool they were asked to use while completing a "search process interview": EDS, Summon, GS, the "conventional" library catalog[18] and research databases, and a "no tool" group that was allowed to choose any search tool to complete the interview. The search process interview took from 30 to 60 minutes and consisted of two parts. In the first section of the interview, students were given four research questions similar to those they might be given for a course assignment and were asked to find two resources

that they would use to complete the assignment (see Appendix A). The first two questions asked for general information about a historical research topic, while the third question asked students to find information to support a sociological argument, and the fourth asked students to find explanatory scientific information and a peer-reviewed scholarly source. These questions generated both observational information about how students approach search tasks and quantitative information about how these tasks were completed. In the second part of the interview, students were asked to reflect on the search strategies they used to complete the test questions and were asked open-ended questions about their search practices and habits, as well as the decision-making processes they used to evaluate resources. These interviews provided qualitative information about how the various search tools fit into students' research workflows.

The search process interview was recorded using the screen capture software Camtasia, which creates a video recording of all information viewable on the monitor during the search session, as well as a synchronized audio recording of the interview participant. The students recorded the URLs for the selected resources in an online data collection form hosted on the Vovici web survey software. The screen capture and audio recordings were transcribed and coded for analysis using the Nvivo qualitative data analysis software, while the quantitative data was analyzed using SPSS statistics software. The interview recordings and transcripts were assigned a code number to ensure confidentiality of the research participants, and the research data were accessible only by members of the research team.

Basic demographic information was collected from all participants including academic year, field of study, number of library instruction sessions attended, and a self-evaluation of the student's ability to locate and evaluate information. Participants were recruited from throughout the IWU and Bucknell undergraduate

	Bucknell		IWU		Total	
	N	% of Bucknell	N	% of IWU	N	% of Total
Natural Sciences & Mathematics	13	28.3%	5	12.2%	18	20.7%
Social Sciences	10	21.7%	8	19.5%	18	20.7%
Humanities	6	13.0%	7	17.1%	13	14.9%
Engineering	7	15.2%	0	.0%	7	8.0%
Visual & Performing Arts	3	6.5%	4	9.8%	7	8.0%
Business, Management, & Accounting	4	8.7%	8	19.5%	12	13.8%
Education	0	.0%	3	7.3%	3	3.4%
Nursing	0	.0%	5	12.2%	5	5.7%
Undeclared	3	6.5%	1	2.4%	4	4.6%
Total	**46**	**100.0%**	**41**	**100.0%**	**87**	**100.0%**

TABLE 1
Participating Students' Fields of Study

population, and represented a variety of disciplines and class years (see table 1).

All students 18 years and older were eligible to participate. In total, 41 students from IWU and 46 students from Bucknell completed the interview (see table 2). Given the breadth of the students who participated, we believe the participants were generally representative of the universities' student populations. However, because all participants were volunteers, this does not constitute a statistically random sample. Participants were recruited through direct e-mails, flyers, ads in the university student newspaper, and posts to an online blackboard system and a university message system. A $20 gift certificate to the bookstore or university café was provided to participants at IWU, while a $10 gift certificate to the university bookstore was provided to students at Bucknell. Both institutions obtained IRB approval for this study, and informed consent was obtained from participants.

To make comparisons between the five test groups, the resources selected by participants were rated by four librarians (two from Bucknell and two from IWU) on a scale from 0 to 3 using a standard rubric developed for the study (see Appendix B). The four librarians independently scored all of the selected resources using the URLs recorded by the participants. While the librarians knew which university a participant was from, they did not know to which test group students were assigned. These scores were used as a relative measure of how well students in each test group completed the research tasks and were analyzed both

TABLE 2
Number of Student Participants in Each Test Group

	Bucknell	IWU	Total
EBSCO Discovery	N/A	11	11
Google Scholar	12	8	20
Library Catalog/Databases	11	14	25
No Tool	11	8	19
Summon	12	N/A	12
Total	**46**	**41**	**87**

in aggregate and on a question-by-question basis (see below).

Agreement between the raters was measured using a Spearman's rho (r_s) correlation for each pair of raters on the scores assigned for each test question (see Appendix C).[19] These correlations were then used to calculate a weighted mean correlation coefficient (see table 3),[20] which indicated that the scores were highly correlated for questions 1–3 as well as the total score, while the fourth question indicated a medium level of correlation.

The consistency of average ratings assigned by the raters was measured using an average measure intraclass correlation coefficient (ICC), which measures the level of agreement between the mean values of the scores given by the four raters. The ICC values indicated that the average ratings were very consistent for questions 1–3 and the total score and moderately consistent for question 4 (see table 3).

Based on these values, we are fairly confident in the reliability of the rater's scores for making comparisons between the test groups. However, it is always prudent to use caution to avoid over-interpreting results based on subjective evaluations. Another potential limitation of this analysis involves the inherent difficulty in comparing students and research tools across universities and, in particular, ascertaining whether students at the two universities were searching a corpus of research materials that is approximately equivalent. For example, the resources available via EDS and Summon depend both on the agreements these services have with other content providers and the subscriptions of the library implementing the discovery tool. While IWU and Bucknell's database subscriptions and physical collections are broadly similar, a systematic comparison of the two libraries' complete holdings and subscriptions was beyond the means of this study. It is therefore possible that the collections differ in some unknown way that could

TABLE 3
Correlation Coefficients for the Mean Score Given to Student Participants on Each Question and in Total

	Weighted Mean Correlation Coefficient	Intraclass Correlation Coefficient
Question 1	0.75	0.900
Question 2	0.71	0.872
Question 3	0.63	0.821
Question 4	0.44	0.738
Total Score	0.73	0.881

Note: All r_s values used to calculate the weighted mean correlation coefficient were significant at $p <$ .01 (one-tailed), and all intraclass correlation coefficient values (ICC (3,4), absolute agreement) fall within a 99% confidence interval.

potentially affect the outcomes of this study. Both EDS and Summon also continuously update their indexes, making it very difficult to precisely determine their overlap, particularly at the level of individual items. Likewise, the scope of GS's index has never been disclosed publicly. While we have attempted to interrogate these potential problems in the following discussion, this limitation could be eliminated in future research by testing multiple discovery tools against a single library collection. Unfortunately, few libraries have implemented multiple discovery tools, and the authors presently know of no studies comparing the usage of these tools in a single-library context.

Quantitative Findings
On the quantitative benchmarks measured by this study, EDS outperformed Summon and the other search systems in almost every category, although not always in a statistically significant way. When evaluated by the librarian raters, the resources located by students using EDS were judged as having higher average quality than any of the other search systems tested (see table 4). The EDS group received the highest total mean score of 2.54, a result that was statistically signifi-

TABLE 4
Means and Standard Deviations of the Scores Obtained by Each Test Group

	EBSCO Discovery		Summon		Google Scholar		Library Catalog/Databases		No Tool	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
All Questions	2.54	.20	1.92	.31	1.80	.23	2.06	.43	2.05	.45
Question 1	2.46	.24	2.29	.56	1.19	.58	2.13	.69	1.96	.68
Question 2	2.20	.43	1.15	.59	1.49	.62	1.94	.81	1.73	.86
Question 3	2.83	.31	2.01	.65	2.33	.50	2.05	.48	2.15	.82
Question 4	2.70	.34	2.19	.55	2.09	.79	2.02	.85	2.33	.42

cant when comparing the EDS test group to all of the other test groups.[21] Students using Summon received a mean score of 1.92, outscoring only the students using GS, who posted the lowest mean score at 1.80. Students using the "conventional" library catalog and databases and the "no tool" groups posted nearly identical mean scores of 2.06 and 2.05 respectively. None of these differences proved statistically significant. At the level of individual research questions, students using EDS posted higher scores that were significant when compared to the GS group on question 1, the Summon group on question 2, the Summon, library catalog/databases, and no-tool groups on question 3, and none of the groups on question 4 (see Appendix D for additional details).

The average scores across the four questions did not vary significantly based on the number of information sessions students reported attending or the level of research skills students felt they had. Students' academic year and academic discipline also did not significantly affect the scores. IWU students' average score was higher than Bucknell students both in aggregate and on every individual question.[22] The effective size of this variance was relatively high, suggesting that the university a particular student attended could explain as much as about 16 percent of the observed variation.[23] However, since the scores within GS, the library catalog/databases, and the no-tool groups did not vary significantly between the two universities, some of this variation may be explained by differences in the functioning of the EDS and Summon tools. Indeed, after removing the scores of the Summon and EDS students from the analysis, the university a student attended explained only about 8 percent of the observed variation.[24]

While not necessarily a measure of research quality, students using EDS required less time to complete the four searches than any of the other test groups (see table 5). However, the range of results between individuals was quite broad, and

TABLE 5
Average Page Views, Searches, and Time Required to Complete the Four Search Tasks for Each Test Group

	Google Scholar			Library Catalog/Databases			No Tool Specified			Summon	EDS
	Bucknell Mean	IWU Mean	Overall Mean	Bucknell Mean	IWU Mean	Overall Mean	Bucknell Mean	IWU Mean	Overall Mean	Bucknell Mean	IWU Mean
Total Page Views	33.4	29.7	31.9	37.6	38.1	37.9	42.8	51.3	46.4	43.5	20.6
Total Number of Searches	9.3	8.23	8.9	11.1	12.5	12.0	13.0	15.8	14.2	9.4	7.4
Total Time to Complete Search Tasks (in Seconds)	987	942	968	885	1,020	963	971	1,232	1,081	1,209	747

the difference in time required to complete the search tasks was not significant when comparing any of the tools.[25] Students using EDS also required fewer searches[26] to find the information they needed and viewed fewer webpages before choosing resources than any of the other four test groups. For page views, these results were significantly different when comparing EDS to all other test groups except the GS group. For the overall number of searches, the results were only significant compared to the no tool group.[27]

Based on these results, it would appear that, in general, EDS was the superior performing discovery system within the parameters of this study, while students using the other three search systems, as well as the students given a choice in search systems, performed more or less equally. The underlying cause of these results, however, bears further scrutiny before making a definitive conclusion. Differences in the types of resources students found and subsequently used might partially explain the observed results. The distribution of resource types used between the test groups was striking, as is shown in table 6.

Just over 20 percent of the resources used by students in the Summon group were non–peer-reviewed newspapers, magazines, and trade journals, compared to only 6.3 percent in the EDS group (and even less in the other three groups). Students in the GS and the no-tool test groups also made frequent use of lower-quality resources. In the GS group, 13 percent of the resources chosen were for-pay articles (usually from Questia or HeinOnline). A little more than 1 in 5 (21.5%) of the resources used by the no-tool group were websites (probably most accurately reflecting students' real-world search habits, as discussed in more detail below). Not surprisingly, books featured much more prominently in the library catalog/databases searches, accounting for 41 percent of the resources chosen, as well as in the GS searches due to its integration with Google Books.

	Google Scholar	Summon	EDS	Library Catalog/ Databases	No Tool
TABLE 6					
Resource Types Chosen by Students in Each Test Group					
Academic Journal Articles	55.0%	65.0%	73.8%	49.2%	50.3%
Books	26.5%	13.4%	12.5%	41.3%	15.4%
Newspapers/Magazines/ Trade Journals	2.0%	20.6%	6.3%	3.2%	2.7%
For-Pay Articles	13.3%	0.0%	0.0%	0.0%	1.3%
Websites (including Wikipedia)	0.7%	0.0%	0.0%	0.0%	21.5%
Government & Legal Document	2.7%	0.0%	5.0%	2.1%	2.0%
Other Documents	0.0%	1.0%	2.5%	4.2%	6.7%

Given that the research questions and scoring rubrics used in this study favored academic books and journal articles (as they likely would on a similar real-world assignment), students who located more of these resources should be likely to obtain higher scores than those who used other resources. In fact, the average scores obtained did loosely correspond with the percentage of students who used books and articles (except for the no-tool group, due to the large number of websites used by students in this group) (see table 7).

Judging from these results, it seems that one of the most important—and perhaps the single most important—factor in determining which resources students will use is the default way in which a particular search system ranks and returns results. For example, the students in the EDS test group may have used fewer newspaper and magazine articles than the Summon group because fewer of these types of articles were ranked highly in the EDS search results. One explanation for this difference may be because IWU's EDS was not set up to search the LexisNexis newspaper database, whereas Bucknell's Summon often returns a great deal of material from this database. However, there is also evidence that Summon's relevancy ranking algorithm ranks newspapers higher than EDS—perhaps even too favorably. In his analysis of Grand Valley State University's implementation of Summon, Doug Way observed a dramatic increase in newspaper usage.[28] Likewise, Bucknell also saw significant increases in newspaper usage after its implementation of Summon at the end of 2009, with yearly usage (measured in click-throughs) of its LexisNexis and

	Google Scholar	Summon	EDS	Library Catalog/ Databases	No Tool
TABLE 7					
Mean Score and Combined Percentage of Book and Journal Article Resources Used by Students in the Test Groups					
Mean Score, All Questions	1.80	1.92	2.54	2.06	2.05
Books & Academic Journal Articles	81.5%	78.4%	86.3%	90.5%	65.8%

ProQuest National newspaper databases increasing over 300 percent and 600 percent respectively (at the end of two years) (see table 8). While Way suggests that Summon might be "meeting untapped demand for aggregated news content," our qualitative observations suggest that Summon might, in fact, be leading students inadvertently to less appropriate resources. The reason for this might be as simple as a small difference in EDS and Summon's relevance ranking. While both systems evaluate content type in their relevancy algorithms, EDS also weights based on article length, meaning that shorter pieces like newspaper articles will rank lower than longer materials like journal articles when other ranking factors are held constant.

Unfortunately, since the relevancy ranking algorithms of both EDS and Summon are proprietary, it is extremely difficult to infer from the search results why particular types of resources are returned. While it appears from our source selection data that there is a systematic difference in the results returned by the EDS instance at IWU and the Summon instance at Bucknell, once newspaper, magazine, and trade periodical results were removed from the scores for the EDS and Summon test groups, the EDS group still obtained higher average scores than the students using Summon

on all four questions. However, the scores for both groups improved (see table 9).[29] Given the multiple variables involved in comparing these two groups, there are several possible interpretations of this result. EDS might be leading students to better resources even beyond the distinction between academic journal articles and books and newspaper, magazine, and trade publication resources. However, the databases included in IWU's installation of EDS may be more suited to these specific research questions than Bucknell's Summon installation. Finally, it is possible that, for whatever reason, the IWU students in the EDS test group were better trained at choosing resources than the students in the Summon test group. Unfortunately, the comparative nature of the data collected in this study prevents a definitive explanation of this issue, and a more detailed study comparing the relevancy rankings of EDS and Summon is probably warranted.

Nevertheless, it seems apparent that setting the default search parameters in Summon to exclude newspaper and magazine articles unless they are specifically queried might help students—albeit unknowingly—use more peer-reviewed academic articles. Allowing libraries to adjust the weighting parameters of the search algorithms themselves might also

TABLE 8 Usage of Selected Newspaper Databases at Bucknell, 2009–2011					
	2009 Click-Throughs	**2010 Click-Throughs**	**Usage Increase Compared to 2009**	**2011 Click-Throughs**	**Usage Increase Compared to 2009**
ProQuest National Newspapers Premier	131	1,475	1026%	918	601%
Ethnic NewsWatch	60	562	837%	481	702%
ABI/INFORM Trade & Industry	28	220	686%	107	282%
America's Historical Newspapers, 1690–1922	15	101	573%	24	60%
LexisNexis Academic	1,280	6,977	445%	5,233	309%
Total, All Databases	49,886	90,854	82%	89,116	79%

TABLE 9
One Way ANOVA Results and Effect Sizes Comparing the Mean Scores of the EDS and Summon Groups, with the Results of Books and Newspapers Excluded

	Mean, EDS	Mean, Summon	F
Question 1	2.46	2.51	$F(1,19) = .098$
Question 2	2.33	1.23	$F(1,15) = 14.66*$
Question 3	2.93	2.28	$F(1,18) = 15.55*$
Question 4	2.7	2.23	$F(1,20) = 5.7*$
Total	2.61	2.18	$F(1,20) = 10.65*$

*Significant at $p < .05$.

help librarians lead students to particular higher-quality resources. None of the search systems tested in this study presently has this functionality. More radically, one could even imagine allowing individual users to adjust and fine-tune search algorithms to reflect their own ranking preferences for a particular search.

Qualitative Findings

While the students using the EDS tool posted higher scores on average than students in the other test groups on the outcomes evaluated in this study, when evaluating students' usage of discovery tools, it is equally important to examine the processes and practices they used to arrive at these outcomes. Throughout the test groups, the researchers observed strong patterns in the way students approached searches no matter which tool they used. These patterns underscore the instructional needs of students in the conceptual aspects of search, in particular the necessity of including algorithmic literacy within a library's information literacy programs.

Students treated almost every search box like a Google search box, using simple keyword searches in 81.5 percent (679/829) of the searches observed (see table 10).[30] This did not vary much by the search tool the student used (although a handful more students using EDS did limit searches by specific criteria). Because of this reliance on simple keyword search, all of the tools tested will typically return a large number of items for a given query. Students were therefore routinely faced with a set of search results that far exceeded what could reasonably be evaluated on an item-by-item basis. This situation of information overabundance makes strategies for evaluating

TABLE 10
Type of Search Conducted by Students in Each Test Group

	Simple Search	Advanced Search Functions (Search Limited by One or More Specific Criteria)	Boolean Search
Google Scholar	94.5%	4.2%	1.4%
Summon	79.3%	12.6%	8.1%
EDS	75.4%	23.1%	1.5%
Library Catalog/ Databases	77.2%	19.1%	3.7%
No Tool	81.1%	16.3%	2.5%
Total, All Groups	81.5%	15.1%	3.4%

and discerning high-quality information of paramount importance. Unfortunately, students often lacked the conceptual understanding required to complete this task adequately, instead relying on the search systems to do the work for them, in particular, by using the search engine's relevancy rankings to determine resources' relative quality.[31]

Students rarely investigated or evaluated sources past the first page of results (see table 11).[32] Fully 92 percent (598/649) of the resources used by students in this study were found on the first page of search results. With the exception of GS, in which more students investigated past the first page, this varied very little between search systems used. By following this practice, students are *de facto* outsourcing much of the evaluation process to the search algorithm itself.

Students' difficulty with understanding and using concepts of how to evaluate search results properly can be compounded by discovery systems that can easily overwhelm a researcher with results they are not equipped to evaluate, sometimes leading students to choose inappropriate resources on which to base their work.[33] For example, trade journals and peer-reviewed journals often appear very similar within organic search results pages.

When evaluating resources, the students interviewed in this study did exhibit an understanding of appropriate methods of ascertaining a source's suitability and quality. During the debriefing interviews, reading abstracts was the most commonly used method of evaluating resources (18% of responses), followed by evaluating a resource's publication location (14%), skim-reading the contents of a resource (10%), and, finally, determining if the resource is peer reviewed (10%). However, more cursory methods—such as simply reading the title—were also very common (12%). Unfortunately, students regularly pursued their evaluation criteria in a superficial way.[34] While students often knew they should look for certain characteristics of a source, they spent very little time and effort actually doing so, instead moving directly to sources. When asked how she decided if a source is reliable, one student exhibited knowledge of ways to evaluate a journal article's potential quality, as well as a willingness to ignore this information, observing, "Generally if it's in a published journal then you are good. More than that, I know with Web of Knowledge you can look and get the impact factor and stuff like that for journals. So you can see if it's a crappy journal or a good journal. I don't usually bother."

Some students also did not fully understand how to define the characteristics of a quality academic article. For example, a sophomore in psychology said that when she evaluated resources, "I always make sure if it's scholarly—if it's supposed to be scholarly—and the years [of the publication] before I start getting any information…because I don't want to waste time getting stuff I don't need." However, when asked, "How would you define a scholarly source?"

TABLE 11		
Percent of Sources Found on the First Page of Search Results for the Five Test Groups		
	Number of Searches Observed	**Percent of Sources Found on First Page**
Google Scholar	138	83%
Summon	91	96%
EDS	77	94%
Library Catalog/Databases	189	94%
No Tool Specified	154	94%

the student admitted, "I don't know the official definition but I think it has to be written by experts in the field, I guess, and maybe reviewed by other scholars. I'm not really sure how it's set up." Students also used superficial judgments when choosing sources. When asked why she had decided against using a particular book as a source, a senior in biology explained, "I'm not sure. I might have eventually used it because I just couldn't find anything else that I wanted. But if I didn't use it, it's because I didn't like the cover very much. Wow, that's a really bad way to pick sources, isn't it?"

Given their uncertainty in evaluating resources, many students imbued the search tools themselves with a great deal of authority.[35] Several students interviewed indicated a high level of trust just in the brand of the search engine or database they used. When asked how she evaluated resources, a first-year student in biomedical engineering said, "I tend to trust in Google Scholar." When asked why, she continued, "Usually the stuff that comes up on this will be published in a major magazine or an online journal. Something that has a reputable standing, other than if you were to search this in Google, it's going to come up with a lot of people's blogs or personal opinions or big threads, whatever it is and you might be able to find a good article."

When asked how he determined if a source was reliable, another student remarked, "It was [in] Google Scholar. It's under my assumption that most things there are scholarly articles that are peer reviewed, researched, cited. Obviously there are a lot of flags you have to look for with general Google. You have to be careful that it's: a) not from Wikipedia; b) not copy-pasted from a pdf. I look at places that don't have citations [with] political agendas, [and avoid places] that don't have authors, [or] don't have biographies on the authors. Generally, if you can get a feel that a writer is reliable and you believe him and he's got citations. It's usually worth [using]."

Likewise, students often place a great deal of trust in the relevancy-ranking algorithms of a particular search engine or database. While discussing how she evaluated the quality of a source, a senior in economics remarked, "Usually the .org or the .edu. And then usually I trust the search engines I'm using too 'cause I trust that [when] I'm using EconLit or JSTOR, the article on there is going to be a scholarly article and not from Wikipedia or something like that."

Many students relied on familiarity with a particular brand for their searches and returned to this search engine or database for research even if it was not the most appropriate or effective. For example, one first-year student in business remarked, "I am familiar with the Academic Search Premier which I use because I've had luck with that in the past. And [Academic Search Premiere] … had more broad [coverage] so that when I search, I feel like I get more things than when I search through some of the other databases titles [that] I wasn't as familiar with…". Another first-year student in biology observed that using Google helped her avoid getting lost in the library's databases: "For something that's kind of general like that, I'd probably go to Google first because it's quicker and you get the results right there and you don't have to worry about is there a full-text online or do I have to order it."

Discovery tools might help eliminate this silo effect of an academic library's diverse databases. A sophomore in biology and classics noted her difficulty choosing which database to use: "I know the database one but sometimes there are some databases where I'm like, I don't know if I should go onto PubMed versus BioOne. I know it has a description but there are so many of them. It's kind of frustrating to go through all of them and find them out. When I was doing my animal behavior paper, I wasn't sure whether to go to the psych articles or the bio articles or the zoology database. I just went to all three of them but it was tedious to do all 3 of them."

Despite the fact that they did not necessarily perform better on the research tasks in this study, students did prefer Google and Google-like discovery tools because they felt that they could get to full-text resources more quickly. The ability to push students to (often lesser-known) full-text resources has been an argument in favor of adopting discovery tools.[36] Students also appear to favor full-text resources and will often avoid requesting articles via interlibrary loan.[37] For example, when describing how she chose a resource, a senior in chemistry explained, "I try to just find something I can open first of all, that has the full text. Then I look to see if it is reliable and it actually has to do with what the thing is…I do this in real life too, I get fed up with searching and it's taking me too long then I'll settle for something else and I'll probably start looking somewhere else."

Discovery systems thus address students' needs by enabling easy cross-database access and access to sources they feel they can trust, especially when compared with Google. When asked what she thought about EDS, a senior in French and international studies explained, "I like it a lot. It's a great starting point to kind of see how many different articles are out there. Sometimes … with Google Scholar…I'd find an article that the abstract sounded nice or the intro sounded like what I was looking for this particular question, but the full text wasn't there. So [EDS] is nice because it can lead me to places where I know I'll have access to the text. Or if not, I can always order it." Similarly, a senior in art history commented: "I think that's why Summon is so good because the results are more than just books and you are able to choose I want scholarly dissertations or I want just journals or newspaper articles. I think when people just search in the catalog, I don't know if they realize that they are just getting books. It sounds kind of dumb, but in order to search newspaper articles or journals, you have to find the specific link where you do that and searching the jour-nals is really hard actually. I would look for specific journals here in the library and the search process, the way it's formatted, is really hard to understand." In this observation, this student encapsulates not only the importance of understanding what resources a discovery tool searches and how these resources are returned and displayed, but also the difficulty in doing so. Unfortunately, too few students understand how these processes and algorithms work, a problem exacerbated by the proprietary designs and complex coverage agreements of the discovery tools themselves.

Conclusions

One of the most powerful features of discovery tools is their ability to meet students' expectations of a single point of entry for their academic research activities supported by a robust and wide-ranging search system. Providing a uniform search interface and aggregating content behind a single "brand," discovery tools like EDS, Summon, and GS helps to diminish the "cognitive load" on students by eliminating the often difficult and confusing step of choosing an appropriate disciplinary database, as well as the need to repeat searches in multiple databases. This might also help simplify user education by allowing instructional librarians to focus on teaching students a single research tool and allowing more time to focus on conceptual research skills, such as evaluating materials.

Not surprisingly, the results of this study have underscored the continued need for research training regardless of the search system implemented. In fact, the relative similarity of the results of students in all of the test groups suggests that well-prepared students can effectively use a variety of search tools, while poorly prepared students will likely struggle even with the best-designed tools. However, the superior performance of the students using EDS also suggests that a particular discovery tool can help lead students to high-quality academic resources. Nevertheless,

as was shown by the relatively lackluster performance in this study of students using Summon, one critical question for libraries considering the implementation of a discovery tool is whether the tool would add enough value to justify its cost in comparison to tools like GS or a library's already implemented suite of research databases. In answering this question, it is especially important to consider not only the quantitative measures of a search tool's efficacy, but also how the search tool fits qualitatively into students' search practices and workflows, and how much a tool contributes positively (or negatively) to a student's overall search experience.

Given a group of search systems—such as those evaluated here—that perform similarly but function differently, the question of which tool or tools to implement and educate students in using becomes one of educational philosophy. Students' practices of primarily using the basic search functionality of any search system, as well as their tendency to rely only on the first page of search results and to trust the relevancy rankings of a given search engine, makes the default settings of these search systems critically important.[38] A careful evaluation should be made of which settings will best serve an institution's students, since these settings will almost certainly have a determinative effect on their research outcomes.

By structuring and ordering the way information is seen and found, any search interface exerts a form of epistemological power by virtue of their relevancy ranking algorithms. The judgments embedded within these systems are often opaque and unclear for the user, but unfortunately they appear to be internalized by many, if not most, students, who routinely trust whatever results a search engine returns.[39] The critical question for librarians is therefore how to participate (or not participate) in this process and what level of this epistemological power to exercise. This is a question that should be explicitly considered by any library that implements a discovery system, as it is clear

that some of the observed deficiencies in students' search practices could be at least partially addressed—without students' knowledge—by choosing to structure the discovery tools' default settings in such a way that students are led to particular types of resources first within the search results. This is by far the most profound difference in the search systems evaluated in this study. Since what is found most quickly and most easily is also what is most likely to be used by students, each system's biases in the types of resources is reflected in the resources they choose.

Future Research

As discovery tools develop and become more popular in use, there exist a number of potentially important avenues for future research. While this study observed undergraduate students searching for preset questions, it would prove useful to study students using discovery tools while conducting their own research for real-life assignments. Examining how these tools perform when used by graduate students, faculty, and librarians while conducting more advanced research may provide further insights into their limitations and benefits. Moreover, it is probable that not all disciplines are equally suited for a discovery tool. For example, Nara Newcomer provides a detailed discussion regarding the specialized information retrieval needs inherent to searching for music materials.[40] Additionally, more in-depth investigations of how particular search tools' relevancy ranking algorithms function and differ is warranted given the critical role they play in how information is presented in the list of results.

Finally, the relationship between discovery tools and information literacy should be evaluated. Does the use of these tools impact what is taught in research instruction sessions? Should the ACRL information literacy standards be rethought in light of these latest tools? In particular, will the ability to evaluate resources become a more highly needed and valued skill?[41] As discovery tools

become more commonplace, librarians will also need to learn how to incorporate them into their research instruction sessions and reference encounters. Will there be an impact on the number and types of instruction sessions requested by faculty, or on the number of reference interactions observed?

Ultimately, discovery tools may, or may not, prove to be the "perfect tool" to compete with Google and keep users engaged with using library resources. However, as discovery tools are adapted and refined, librarians must be involved in assessing their effectiveness, impact, and usability.

APPENDIX A.

Research Questions Given to Students Participating in This Study

1. You need to give a class presentation that explains general information about the Civil Rights Act of 1964.
    Find 2 sources that you would use as the basis of your presentation.

2. You need to find information about women's professional baseball in the 1940s.
    Find 2 sources that would give you this information.

3. You are writing a research paper that argues that increased wealth does not result in increased happiness.
    Find 2 of the best-quality sources to use.

4. You are writing a research paper on how volcanic eruptions affect the Earth's climate.
    Your professor has told you to use only peer-reviewed, scholarly articles.
    Find 2 sources that you might use.

APPENDIX B.

Scoring rubric used in this study:

The resource scores 3 if:

- It provides a sufficient overview of the topic appropriate to use as the basis for an academic presentation assignment.
- The source directly addresses the research question.
- Source is sufficiently reliable for use in an academic setting.
- The resources are appropriately up-to-date for the research question.
- Source is not drawn from a primary text that lacks adequate background information.

Question-specific requirements for a score of 3:
- For question 1: The resource must not be too detailed for a general presentation.
- For question 2: The resource should not be a primary text about women's baseball (e.g. newspaper articles from the time period, obituaries of players, etc.).
- For question 3: Source must provide reliable data on which to base the argument that wealth does not increase happiness.
- For question 4: The source must be a peer-reviewed scholarly work.

Typical examples: Journal articles, academic books, secondary or tertiary sources that have been adequately reviewed, scholarly reference works, websites of high academic quality.

The resource scores 2 if:
- Source is likely reliable, but is deficient in no more than one of the criteria required to score a 3, such as:
  ◦ Materials are out-of-date.
  ◦ Materials do not provide sufficient context.
  ◦ Articles that don't directly address the research question.

Typical examples: Journal articles that are too highly specialized, high-quality magazine articles, general audience books, legal texts and articles, many primary texts, websites of good quality, for-pay articles that are high quality and with a free option.

The resource scores 1 if:
- Source is of questionable reliability and/or authority.
- The source is deficient in multiple criteria required to score a 3.

Typical examples: Newspaper articles, trade magazines, lower-quality webpages, most Wikipedia articles, low-quality journal articles and books, for-pay articles that are likely to be of good quality.

The resource scores 0 if:
- Student fails to complete the task.
- The source is not relevant to the question topic.
- Source is unlikely to be acceptable in a classroom.

Typical examples: Any resources listed above that are of minimal academic value, for-pay articles that are likely to be poor quality.

APPENDIX C.

The following tables indicate the Spearman's rho (rs) correlations for each pair of raters on the scores assigned for each test question and on all questions combined, as well as the number of score pairs (N) used to calculate the correlation. The weighted mean correlation coefficient is given at the end of each table. The number of score pairs included in each correlation (as well as in the ICC calculation above) differ between the pairs of raters due to technical issues that prevented all raters from scoring every resource provided by the students. These technical problems were nonsystematic, and we do not believe that they affect the values of these calculations in a significant way.

Question 1

		Rater 1	Rater 2	Rater 3	Rater 4
Rater 1	r_s		0.715	0.786	0.72
	N		81	84	81
Rater 2	r_s	0.715		0.747	0.72
	N	81		83	79
Rater 3	r_s	0.786	0.747		0.775
	N	84	83		82
Rater 4	r_s	0.720	0.720	0.775	
	N	81	79	82	
Weighted Mean Correlation Coefficient = 0.746					

Question 2

		Rater 1	Rater 2	Rater 3	Rater 4
Rater 1	r_s		0.729	0.762	0.752
	N		72	78	77
Rater 2	r_s	0.729		0.693	0.602
	N	72		74	73
Rater 3	r_s	0.762	0.693		0.709
	N	78	74		81
Rater 4	r_s	0.752	0.602	0.709	
	N	77	73	81	
Weighted Mean Correlation Coefficient = 0.712					

Question 3

		Rater 1	Rater 2	Rater 3	Rater 4
Rater 1	r_s		0.639	0.69	0.728
	N		74	83	81
Rater2	r_s	0.639		0.512	0.482
	N	74		77	73
Rater 3	r_s	0.690	0.512		0.687
	N	83	77		82
Rater 4	r_s	0.728	0.482	0.687	
	N	81	73	82	
Weighted Mean Correlation Coefficient = 0.635					

Question 4

		Rater 1	Rater 2	Rater 3	Rater 4
Rater 1	r_s		0.461	0.661	0.463
	N		69	76	77
Rater 2	r_s	0.461		0.308	0.368
	N	69		74	74
Rater 3	r_s	0.661	0.308		0.348
	N	76	74		83
Rater 4	r_s	0.463	0.368	0.348	
	N	77	74	83	
Weighted Mean Correlation Coefficient = 0.443					

All Questions

		Rater 1	Rater 2	Rater 3	Rater 4
Rater 1	r_s		0.733	0.791	0.678
	N		56	70	75
Rater 2	r_s	0.733		0.786	0.621
	N	56		51	52
Rater 3	r_s	0.791	0.786		0.711
	N	70	51		67
Rater 4	r_s	0.678	0.621	0.711	
	N	75	52	67	
Weighted Mean Correlation Coefficient = 0.725					

APPENDIX D.

The following tables give the results for one-way ANOVA and Tukey post-hoc tests used to compare the mean scores of the test groups for all scores combined and for each test question.

All Scores Combined
One-Way ANOVA Result: $F(4,81) = 7.416$, $p = .000$.
T**ukey Post-Hoc Analysis Results:**

Group:	Comparison Group:	Mean Difference (I–J)	Std. Error	p
EBSCO Discovery	Summon	.62645*	.15438	.001
	Library Catalog/Databases	.48778*	.13490	.005
	Google Scholar	.74162*	.13964	.000
	No Tool Specified	.49080*	.14086	.007
Summon	EBSCO Discovery	−.62645*	.15438	.001
	Library Catalog/Databases	−.13867	.12662	.809
	Google Scholar	.11517	.13165	.905
	No Tool Specified	−.13565	.13294	.845
Library Catalog/ Databases	EBSCO Discovery	−.48778*	.13490	.005
	Summon	.13867	.12662	.809
	Google Scholar	.25384	.10816	.141
	No Tool Specified	.00303	.10973	1.000

Google Scholar	EBSCO Discovery	−.74162*	.13964	.000
	Summon	−.11517	.13165	.905
	Library Catalog/Databases	−.25384	.10816	.141
	No Tool Specified	−.25081	.11550	.201
No Tool Specified	EBSCO Discovery	−.49080*	.14086	.007
	Summon	.13565	.13294	.845
	Library Catalog/Databases	−.00303	.10973	1.000
	Google Scholar	.25081	.11550	.201
* The mean difference is significant at the 0.05 level.				

Question 1

One-Way ANOVA Result: $F_{(4,81)} = 11.063$, $p = .000$.
Tukey Post-Hoc Analysis Results:

Group:	Comparison Group:	Mean Difference (I–J)	Std. Error	Sig.
EBSCO Discovery	Summon	.17530	.26094	.962
	Library Catalog/Databases	.32964	.22803	.600
	Google Scholar	1.27500*	.23603	.000
	No Tool Specified	.50385	.23809	.223
Summon	EBSCO Discovery	−.17530	.26094	.962
	Library Catalog/Databases	.15435	.21403	.951
	Google Scholar	1.09970*	.22253	.000
	No Tool Specified	.32856	.22472	.590
Library Catalog/ Databases	EBSCO Discovery	−.32964	.22803	.600
	Summon	−.15435	.21403	.951
	Google Scholar	.94536*	.18283	.000
	No Tool Specified	.17421	.18548	.881
Google Scholar	EBSCO Discovery	−1.27500*	.23603	.000
	Summon	−1.09970*	.22253	.000
	Library Catalog/Databases	−.94536*	.18283	.000
	No Tool Specified	−.77115*	.19524	.002
No Tool Specified	EBSCO Discovery	−.50385	.23809	.223
	Summon	−.32856	.22472	.590
	Library Catalog/Databases	−.17421	.18548	.881
	Google Scholar	.77115*	.19524	.002
* The mean difference is significant at the 0.05 level.				

Question 2
One-Way ANOVA Result: $F_{(4, 80)}$ = 4.124, p = .004.
Tukey Post-Hoc Analysis Results:

Group:	Comparison Group:	Mean Difference (I–J)	Std. Error	Sig.
EBSCO Discovery	Summon	1.04524*	.30588	.009
	Library Catalog/Databases	.25635	.26888	.875
	Google Scholar	.71339	.27668	.084
	No Tool Specified	.46692	.27910	.456
Summon	EBSCO Discovery	−1.04524*	.30588	.009
	Library Catalog/Databases	−.78889*	.25257	.020
	Google Scholar	−.33185	.26086	.709
	No Tool Specified	−.57832	.26342	.192
Library Catalog/ Databases	EBSCO Discovery	−.25635	.26888	.875
	Summon	.78889*	.25257	.020
	Google Scholar	.45704	.21629	.225
	No Tool Specified	.21057	.21937	.872
Google Scholar	EBSCO Discovery	−.71339	.27668	.084
	Summon	.33185	.26086	.709
	Library Catalog/Databases	−.45704	.21629	.225
	No Tool Specified	−.24648	.22886	.818
No Tool Specified	EBSCO Discovery	−.46692	.27910	.456
	Summon	.57832	.26342	.192
	Library Catalog/Databases	−.21057	.21937	.872
	Google Scholar	.24648	.22886	.818
*The mean difference is significant at the 0.05 level.				

Question 3
One-Way ANOVA Result: $F_{(4,81)} = 3.766$, $p = .007$.
Tukey Post-Hoc Analysis Results:

Group:	Comparison Group:	Mean Difference (I–J)	Std. Error	Sig.
EBSCO Discovery	Summon	.81310*	.25131	.015
	Library Catalog/Databases	.77038*	.21961	.006
	Google Scholar	.50000	.22731	.190
	No Tool Specified	.67932*	.22930	.032
Summon	EBSCO Discovery	−.81310*	.25131	.015
	Library Catalog/Databases	-.04271	.20612	1.000
	Google Scholar	−.31310	.21431	.591
	No Tool Specified	−.13377	.21642	.972
Library Catalog/ Databases	EBSCO Discovery	−.77038*	.21961	.006
	Summon	.04271	.20612	1.000
	Google Scholar	−.27038	.17608	.543
	No Tool Specified	−.09106	.17863	.986
Google Scholar	EBSCO Discovery	−.50000	.22731	.190
	Summon	.31310	.21431	.591
	Library Catalog/Databases	.27038	.17608	.543
	No Tool Specified	.17932	.18803	.875
No Tool Specified	EBSCO Discovery	−.67932*	.22930	.032
	Summon	.13377	.21642	.972
	Library Catalog/Databases	.09106	.17863	.986
	Google Scholar	−.17932	.18803	.875
* The mean difference is significant at the 0.05 level.				

Question 4

One-Way ANOVA Result: $F(4,81) = 2.099$, $p = .088$.

Tukey Post-Hoc Analysis Results:

Group:	Comparison Group:	Mean Difference (I–J)	Std. Error	Sig.
EBSCO Discovery	Summon	.50952	.28864	.401
	Library Catalog/Databases	.67524	.25223	.066
	Google Scholar	.60536	.26108	.150
	No Tool Specified	.36729	.26336	.633
Summon	EBSCO Discovery	−.50952	.28864	.401
	Library Catalog/Databases	.16571	.23674	.956
	Google Scholar	.09583	.24615	.995
	No Tool Specified	−.14223	.24857	.979
Library Catalog/Databases	EBSCO Discovery	−.67524	.25223	.066
	Summon	−.16571	.23674	.956
	Google Scholar	−.06988	.20223	.997
	No Tool Specified	−.30794	.20517	.565
Google Scholar	EBSCO Discovery	−.60536	.26108	.150
	Summon	−.09583	.24615	.995
	Library Catalog/Databases	.06988	.20223	.997
	No Tool Specified	−.23806	.21596	.805
No Tool Specified	EBSCO Discovery	−.36729	.26336	.633
	Summon	.14223	.24857	.979
	Library Catalog/Databases	.30794	.20517	.565
	Google Scholar	.23806	.21596	.805

Notes

1. For a thorough discussion on research behaviors in the digital age, see various Project Information Literacy reports, available online at http://projectinfolit.org/publications/; also Ian Rowlands et al., "The Google Generation: The Information Behaviour of the Researcher of the Future," *Aslib Proceedings* 60, no. 4 (2008): 290–310.

2. For a discussion on the issue of "convenience" when seeking information and the relationship between libraries and Google, see Lynn Silipigni Connaway, Timothy J. Dickey, and Marie L. Radford, "'If It Is Too Inconvenient I'm Not Going After It': Convenience as a Critical Factor in Information-Seeking Behaviors," *Library & Information Science Research* 33, no. 3 (July 2011): 179–90.

3. Illinois Wesleyan University and Bucknell University are both private, highly selective liberal arts universities with 2,100 and 3,600 students respectively.

4. Judy Luther, "Trumping Google? Metasearching's Promise," *Library Journal* 128, no. 16 (Oct. 10, 2003): 36–39.

5. Abe Korah and Erin Dorris Cassidy, "Students and Federated Searching: A Survey of Use and Satisfaction," *Reference & User Services Quarterly* 49, no. 4 (July 15, 2010): 325–32.

6. Jeff Wisniewski, "Web Scale Discovery: The Future's So Bright, I Gotta Wear Shades," *Online* 34, no. 4 (July 2010): 55–57; Doug Way, "The Impact of Web-Scale Discovery on the Use of a Library Collection," *Serials Review* 36, no. 4 (Dec. 2010): 214–20.

7. "About Google Scholar," available online at http://scholar.google.com/intl/en/ scholar/ about.html [accessed 5 January 2012].

8. Gail Herrera, "Google Scholar Users and User Behaviors: An Exploratory Study," *College & Research Libraries* 72, no. 4 (July 2011): 316–31; Amy Hoseth, "Google Scholar," *Charleston Advisor* 13, no. 1 (Jan. 2011): 36–39.

9. Rebecca Donlan and Rachel Cooke, "Running with the Devil: Accessing Library-Licensed Full-Text Holdings Through Google Scholar," *Internet Reference Services Quarterly* 10, no. 3/4 (July 2005): 149–57. See also Bonnie Imler and Michelle Eichelberger, "Do They 'Get It'? Student Usage of SFX Citation Linking Software," *College & Research Libraries* 72, no. 5 (Sept. 2011): 454–63 for a discussion of student usage of SFX citation linking software.

10. For a thorough discussion of web scale discovery tools, see Jason Vaughan, "Chapter 1: Web Scale Discovery What and Why?" *Library Technology Reports* 47, no. 1 (Jan. 2011): 5–11; Jason Vaughan, "Chapter 6: Differentiators and A Final Note," *Library Technology Reports* 47, no. 1 (Jan. 2011): 48–53.

11. Josh Hadro, "Competition Heats Up Discovery Marketplace," *Library Journal* 135, no. 17 (Oct. 15, 2010): 14; David Aymonin et al., "Be Realistic, Demand the Impossible: Comparison of 4 Discovery Tools Using Real Data at the EPFL Library," *Technical Report* (Dec. 19, 2011): 1–32, available online at http://infoscience.epfl.ch/record/172947 [accessed 15 January 2012].

12. Way, "The Impact of Web-Scale Discovery," 214–16.

13. For a usability study focusing on AquaBrowser, Encore, Primo, and VuFind, see Karen Joc and Kayo Change, "The Impact of Discovery Platforms on the Information-Seeking Behaviour of EFL Undergraduate Students," *VALA2010 Conference* (n.d.): 1–22, available online at *www.vala. org.au/vala2010/.../VALA2010_122_Joc_Final.pdf [need access date].*

14. Julia Gross and Lutie Sheridan, "Web Scale Discovery: The User Experience," *New Library World* 112, no. 5/6 (June 2011): 236–47.

15. Sarah C. Williams and Anita K. Foster, "Promise Fulfilled? An EBSCO Discovery Service Usability Study," *Journal of Web Librarianship* 5, no. 3 (Sept. 2011): 179–98.

16. David Howard and Constance Wiebrands, "Culture Shock: Librarians' Response to Web Scale Search," *Conference Proceedings, ALIA Information Online Conference* (Feb. 2011), available online at http://ro.ecu.edu.au/ecuworks/6206/ [accessed 28 December 2011].

17. See Noah Brubaker, Susan Leach-Murray, and Sherri Parker, "Shapes in the Cloud: Finding the Right Discovery Layer," *Online* 35, no. 2 (Mar. 2011): 20–26; Ronda Rowe, "Web-Scale Discovery: A Review of Summon, EBSCO Discovery Service; and WorldCat Local," *Charleston Advisor* 12, no. 1 (July 2010): 5–10.

18. Sirsi at Bucknell and the Voyager VuFind interface at IWU.

19. Spearman's rho was chosen as the most appropriate measure of interrater reliability for this study because questions were evaluated by multiple raters on a 0–3 ordinal scale. For additional information on the use of this statistic, see Philip Bobko, *Correlation and Regression: Applications for Industrial Organizational Psychology and Management*, 2nd ed. (London: Sage, 2001): 31–33.

20. A specific procedure is required to calculate the weighted mean correlation coefficient. First, the Spearman's rho correlation values for each pair of raters must be converted to Fisher's z values using a Fischer's z transformation. These z values are then used to calculate a weighted average for all values that takes into account the sample size for each pair of raters. Finally, this weighted average of z values is "back-converted" using an inverse Fisher's transformation to produce an approximate weighted mean correlation coefficient. See Bobko, *Correlation and Regression*, 48–53, for a detailed explanation and examples of this procedure.

21. One-way ANOVA tests were conducted to compare the mean scores of the students in each test group both in total (using the scores for all eight resources obtained) and on each individual question (two scores per question). All of these tests indicated significant differences among the groups (at $p < .05$), and Tukey post-hoc tests were used to determine the significance of differences

between specific groups (for detailed tables, see Appendix D).

22. This result was significant at $p < .05$ using a using a one-way ANOVA for all questions combined and questions 1–3, but not significant on question 4. The full ANOVA results were as follows: All questions combined: $F(1,84) = 15.831$, $p = .000$, eta squared = .159; Question 1: $F(1,84) = 7.933$, $p = .006$, eta squared = .086; Question 2: $F(1,83) = 7.611$, $p = .007$, eta squared = .084; Question 3: $F(1,84) = 5.235$, $p = .025$, eta squared = .059; Question 4: $F(1,84) = 0.420$ $p = .519$. eta squared = .005.

23. The eta squared value was .159 for all questions combined. See note 24.

24. The eta squared value for the variance between universities decreased to 0.82 using a one-way ANOVA ($F(1,62) = 5.557$, $p = .022$).

25. Using a one-way ANOVA at $p < .05$.

26. For the purposes of this study, a new search was defined as whenever the student entered a new set of terms into a search interface and produced a search result.

27. Using a one-way ANOVA and Tukey post-hoc tests at $p < .05$.

28. Way, "The Impact of Web-Scale Discovery," 217–18.

29. These differences were statistically significant in 3 out of 4 questions using a one-way ANOVA at $p < .05$.

30. See also Andrew D. Asher and Lynda M. Duke, "Searching for Answers: Student Research Behavior at Illinois Wesleyan University," in *College Libraries and Student Culture: What We Now Know*, eds. Lynda M. Duke and Andrew D. Asher (Chicago: American Library Association, 2012):77; S. Hampton-Reeves et al., *Students' Use of Research Content in Teaching and Learning: A Report for the Joint Information Systems Council (JISC)* (Centre for Research-Informed Teaching, University of Central Lancashire, 2009),45; CIBER (Centre for Information Behaviour and Evaluation of Research), *Information Behaviour of the Researcher of the Future: A CIBER Briefing Paper* (London: CIBER, 2008),14.

31. See also Judit Bar-Ilan et al., "Presentation Bias Is Significant in Determining User Preference for Search Results: A User Study," *Journal of the American Society for Information Science and Technology* 60, no. 1 (Jan. 1, 2009); Bing Pan et al., "In Google We Trust: Users' Decisions on Rank, Position, and Relevance," *Journal of Computer-Mediated Communication* 12, no. 3 (Apr. 2007): 816; L.A. Granka, T. Joachims, and G. Gay, "Eye-Tracking Analysis of User Behavior in WWW Search," in *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (2004), 478–79.

32. See also Asher and Duke, "Searching for Answers," 80; Bernard J. Jansen and Amanda Spink, "How Are We Searching the World Wide Web? A Comparison of Nine Search Engine Transaction Logs," *Information Processing & Management* 42, no. 1 (Jan. 2006): 257–58; J.R. Griffiths and P. Brophy, "Student Searching Behavior and the Web: Use of Academic Resources and Google," *Trends* 53, no. 4 (2005): 551; CIBER, "Information Behaviour," 10.

33. Asher and Duke, "Searching for Answers," 80–81; see also H.L. Lee, "Information Structures and Undergraduate Students," *Journal of Academic Librarianship* 34, no. 3 (2008): 215.

34. See also Asher and Duke, "Searching for Answers," 80–82.

35. See also B.J. Jansen, M. Zhang, and C.D. Schultz, "Brand and Its Effect on User Perception of Search Engine Performance," *Journal of the American Society for Information Science and Technology* 60, no. 8 (2009); E. Hargittai, et al., "Trust Online: Young Adults' Evaluation of Web Content," *International Journal of Communication* 4 (2010).

36. Way, "The Impact of Web-Scale Discovery," 219.

37. See also Lynn Silipigni Connaway and Thomas Dickey, "The Digital Information Seeker: Report on Findings from Selected OCLC, RIN and JISC User Behaviour Projects" (OCLC Research, 2010), 27.

38. Siva Vaidhyanathan, *The Googlization of Everything (and Why We Should Worry)* (Berkeley: University of California Press, 2011), 88–90.

39. Andrew Asher, "Search Magic: Discovering How Undergraduates Find Information" (paper presented at the American Anthropological Association Annual Meeting, Montreal, Canada, Nov. 18, 2011), available online at www.andrewasher.net/anthropologyofalgorithms/?p=5 [need access date]; Ted Striphas, "Who Speaks for Culture?" available online at www.thelateageofprint.org/2011/09/26/who-speaks-for-culture/ [accessed 23 July 2013]; Ted Striphas, "Culturomics," available online at www.thelateageofprint.org/2011/04/05/culturomics/ [accessed 23 July 2013].

40. Nara L. Newcomer, "The Detail Behind Web-Scale: Selecting and Configuring Web-Scale Discovery Tools to Meet Music Information Retrieval Needs," *Music Reference Services Quarterly* 14, no. 3 (2011): 131–45.

41. For an interesting discussion of discovery tools and information literacy standards, see Jody Condit Fagan, "Discovery Tools and Information Literacy," *Journal of Web Librarianship* 5, no. 3 (2011): 171–78; David Howard and Constance Wiebrands, "Culture Shock: Librarians' Response to Web Scale Search," *Conference Proceedings, ALIA Information Online Conference,* (Feb. 2011), available online at http://ro.ecu.edu.au/ecuworks/6206/ [accessed 28 December 2011].