7-1-2011

Google Scholar Users and User Behaviors: An Exploratory Study

Gail Herrera
University of Mississippi, gherrera@olemiss.edu

Google Scholar Users & User Behaviors: An Exploratory Study

Gail Herrera
Assistant Dean, Technical Services & Automation
Associate Professor
University of Mississippi Libraries
e-mail: gherrera@olemiss.edu

Google Scholar Users & User Behaviors: An Exploratory Study

ABSTRACT.
The University of Mississippi Library created a profile to provide linking from Google Scholar to library resources in 2005. Although Google Scholar does not provide usage statistics for institutions, use of Google Scholar is clearly evident in looking at library link resolver logs. The purpose of this project is to examine users of Google Scholar using existing data from interlibrary loan transactions and library website click-through logs and analytics. Questions about user status and discipline as well as behaviors related to use of other library resources, are explored.

The University of Mississippi is one of three major higher education universities in the state with

12,762 undergraduate students, 1,865 graduate students and 714 faculty.[i] In 2005, the University

of Mississippi Library began participating in the Google Scholar Library Links program. This

program allows users associated with an institution to access their library's link resolver and

associated full text. For the University of Mississippi, this means that users with a campus IP

address are automatically assigned our institution's profile and link resolver. Off-campus users

have to select the University of Mississippi in their Scholar Preferences before searching Google

Scholar in order to obtain these same features.

Figure 1. Link Resolver Clicks by Source
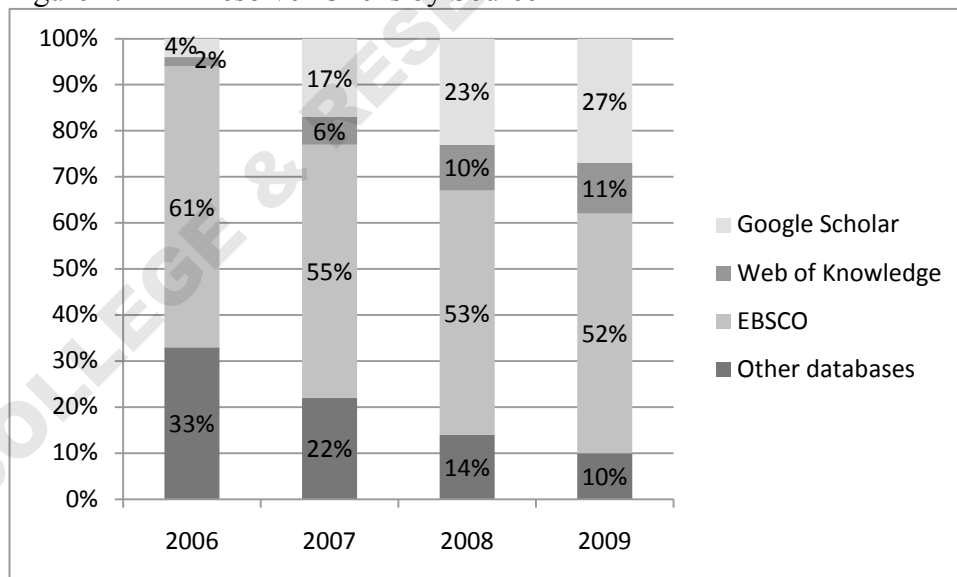
Having implemented the Google Scholar Library Links profile late in 2005, the first full year of data began in 2006. Figure 1 depicts the total percentage of clicks from their source through our link resolver for 2006 to 2009. Sources were broken down into four headings: Google Scholar, Web of Knowledge, EBSCO and other databases. The University of Mississippi Library subscribes to many EBSCO databases (see Appendix A) that are viewed here as one source. Web of Knowledge includes Science Citation Index, Social Sciences Citation Index and Medline. The "other databases" category is a single group that consists of all other databases that make up very small percentages. Over the past four years, the percentage of clicks coming from Google Scholar (GS) has steadily increased from four percent in 2006 to twenty-seven percent in 2009. The total percentage of clicks from Web of Knowledge has also steadily increased from two percent to eleven percent. Although the percentage of clicks coming from EBSCO databases has lessened, it still is the predominant source with GS ranking second.

Literature Review

Google Scholar has been a substantial research topic over the past five years. Early on, research focused primarily on functionality and content. Since GS's entrance to the library world with its library links program in 2005, interest in this perpetual beta product has grown. One of the tensions between libraries and GS has been the need for transparency. Although Google should be applauded for their library links program, which is a very positive service, they have not been open about their content partners, content updating and coverage,[ii] nor their relevancy algorithms related to "citedness."[iii] There have also been concerns related to functionality including searching, linking, indexing, quality control,[iv] the lack of sorting features, and compatibility with citation management software.[v] The familiar and easy search tool utilized by GS is no doubt attractive and requires little to no instruction unlike other citation tools such as Web of

Science/Knowledge. In an interview, GS's founding engineer, Anurag Acharya, described usage as growing exponentially.[vi] GS usage information is not available to participating institutions or libraries.

In relation to content, there have been changes over the past five years. Hartman and Mullen noted in their 2007 GS update that Elsevier (ScienceDirect) had not been a part of the publisher partner program in 2005, but their subsequent addition to the program had increased the amount of linked content.[vii] According to Vine, the currency of PubMed citations in GS has improved from being more than a year behind to being several months behind.[viii] In 2005, Noruzi noted that GS did not index articles in Persian or Chinese.[ix] Language bias was also noted in other studies as well.[x] In 2008, however, Acharya discussed GS's plans to incorporate "significant coverage in Chinese, German, French, Portuguese, Spanish, Japanese and soon Korean."[xi]

New metrics for calculating journal impact have also been proposed since the advent of GS.[xii] Meho and Yang compared Web of Science with Scopus and Google Scholar using citations from library and information science faculty members. They found that use of all three tools provide the most comprehensive picture of an author's overall impact. They noted GS's coverage of conference proceedings as well as international, non-English language journals but also noted the significant amount of time it required to analyze GS results.[xiii] According to Hartman and Mullen, GS has "gained popularity as a free and effective alternative to Web of Science and Scopus, the more traditional subscription citation analysis tools found in most academic libraries."[xiv]

Several studies have compared GS to other library databases. Gardner and Eng's comparison to ERIC, PsycINFO, and SSCI noted a greater variety of sources and citations

coming from GS but a lack of coverage for more recent literature. Neuhaus et al. compared

content of forty-seven databases and GS using random samples. They found that content covered

by GS varied. Disciplinary content strengths were in the sciences and medicine. Weaknesses

were identified in the social sciences and humanities.[xv] Callicott and Vaughn found in

comparing GS to a library catalog and EBSCO's Academic Search Premier that GS would be

best positioned as a supplementary research tool and noted the relevance of GS results in the

humanities to be "surprisingly solid."[xvi] Schroeder's 2007 finding in reviewing ten GS and Web

of Science studies noted the inclusion of valued materials such as conference proceedings,

books, preprints, institutional repository information and open access content in GS not found in

Web of Science. She also pointed out the issue of GS "false hits" that require the user to spend

more time analyzing results.[xvii]

In 2007, Christianson evaluated the indexing of 840 articles from core ecology journals.

Only fifty-seven percent of the test articles were included in GS with full citations while seventy-

seven percent were included with at least partial citations. Based on findings, Christianson

pointed out the validity of libraries directing users to their licensed databases while

acknowledging that GS often provides "good-enough" results.[xviii] Levine-Clark and Kraus

compared GS with Chemical Abstracts Service (CAS) and found that GS returned more results

for topical searches. However, CAS returned more results for chemical compound and personal

name searches.[xix] Meier and Conkling compared GS with Compendex in engineering. They

concluded that GS is a useful tool for engineering literature covering the past ten to fifteen years.

Further, the authors identified GS as a good starting place for undergraduate research projects.[xx]

In the past year, there have been even more comparisons with databases that perhaps note changes to the earlier disciplinary weaknesses. For example, Walters compared search performance in the subject area of later-life migration in GS and eleven other library subscription databases including EBSCO's Academic Search Elite, MEDLINE, SSCI and SocINDEX. He found that in both recall and precision, GS performed "better than most of the subscription databases."[xxi] Howland et al. set out to compare the scholarliness of GS with traditional library resources. Within broad academic disciplines, they matched search terms in specific disciplines with traditional library databases and compared them with GS. On average, they found GS content to be more scholarly than library databases and further concluded there to be no statistically significant differences in the level of scholarliness across disciplines. In fact, GS included seventy-six percent of all the citations found in library databases. Conversely, the library databases contained only forty-seven percent of the GS citations. [xxii] Schroeder,[xxiii] Gardner and Eng,[xxiv] Howland et al.,[xxv] and Meho and Yang[xxvi] all noted a greater variety of resources and more results in GS in comparison to traditional library databases.

Adoption of GS among libraries has also been a research topic. Mullen and Hartman, in their 2007 analysis of ARL library websites, found that a "significant association between partner status and number of paths to Google Scholar" had developed since their original study in 2005 and further declared GS to have found a place as a discovery tool in ARL libraries.[xxvii] Neuhaus et al. reviewed the adoption of GS on college and university websites. They found that seventy-three percent of the research institutions studied were providing access from GS to their link-resolved, library resource.[xxviii]

Researchers have compared Google Scholar with federated or metasearch tools. One usability study observed thirty-two undergraduates in their use of Google Scholar and MetaLib.

They determined that GS "performed better in almost all measures" and found students more positive about GS.[xxix] Xiaotian's comparison concluded that MetaLib and WebFeat could not "compete with Google in speed, simplicity, ease of use, and convenience…."[xxx] GS's ability to reveal open access materials has also been noted as a benefit over library-developed search engines.[xxxi] The concluding themes of most research validated the usefulness of adding GS as a complementary research tool.

A number of user studies also shed some light on the use of Google among students and researchers. Many of the studies demonstrate the importance of search engines like Google for today's students and researchers. For example, a study on student searching behavior by Griffiths and Brophy found that forty-five percent of students used Google as their "first port of call when locating information."[xxxii] OCLC's 2005 user study on college students found that eighty-nine percent of college students use search engines to begin an information search while only two percent use the library website for this same purpose.[xxxiii] They also found that sixty-eight percent of college students indicated that Google was the search engine they had most recently used.[xxxiv] Evidence along these same lines exists for researchers.

An observational study of researchers in Stockholm, Sweden, was conducted from 2005-2006 that revealed researchers "used Google for everything"[xxxv] and were surprised by the "almost complete dominance of Google as a starting point for searching scientific information…." They concluded that for many researchers, and especially for those in the sciences, "Google is the first choice for information – all kinds of information."[xxxvi] In fact, some researchers even stated having moved away from "subject specific databases to Google (and Google Scholar)."[xxxvii]

In the Ithaka 2009 Faculty Survey, discovery through Google and Google Scholar came in third place for faculty when asked how often they used different methods to find information in academic journals.[xxxviii] They also found digital versions to be the preferred format for most faculty members."[xxxix] Another study surveyed 2,063 academic researchers in natural science, engineering, and medical science from five research universities in the United States. They concluded that "differences in information-seeking behavior among universities are not as clear as among disciplines and demographics."[xl] All of these user studies show a general adoption of Google among students and researchers alike.

Methods

The purpose of this research was to analyze existing logs to explore user information such as broad discipline and status along with use of other library resources. This exploratory study of GS use and users builds upon existing research. It looks at the adoption of GS at a user level rather than an institutional or library level and explores several questions:

- Which disciplines are using GS?

- Are there statistically significant differences between GS users and non-users?

- What types of users (undergraduates, graduates, and faculty) are using GS?

- Are GS users using other library resources?

Although the link resolver shows high use, link resolver data is not tied to user information such as discipline or user status. In order to obtain samples of GS users, two approaches were taken. First, interlibrary loan requests were analyzed. Second, use of GS from library website click-throughs were examined. This part of the project required combining two different logs that match the IP address with a patron record number. The patron record number then served as a lookup matchpoint in the data when combined with a patron database export that included the

patron record number, barcode and department affiliation. Most systems provide summary reports related to user status and department. Since this process combined logs from two different systems, a lookup was created using Microsoft Access.

Both of these user samples were small, especially when compared to the overall campus population. The library's fund hierarchy is divided into three broad disciplines including the humanities, social sciences and sciences. Three Senior Subject Librarians lead these disciplines with more than a dozen Subject Librarians who serve as liaisons and instructors to the departments in these areas. These disciplines are used throughout this study for local resource allocation purposes. Evidence in the literature also suggests "research practices and teaching methods have both shifted, most often at a disciplinary level."[xli] Google Analytics, a tool for gathering website usage information, also provided information about keywords leading GS users to library resources. Using SPSS, statistical significance ($p<.05$) was determined for detecting statistically significant parterns. These research methods were reviewed and approved by the University's Institutional Review Board.

Targets

First, it is important to review the targets available through the library's link resolver to ensure that journal titles are available in all three disciplines. For 2009, GS users linked to ninety-six different targets. GS covers more targets than any other source, including Web of Knowledge and EBSCO. In the number of targets covered, it is only comparable to the library's many EBSCO databases combined. Two of the ninety-six targets were services including the library catalog and interlibrary loan.

In 2009, there were 1,290 clicks coming from GS to the library catalog. In calculating the overall percent of GS clicks, however, only twelve percent went to the library catalog. By

comparison, fifty percent of Web of Knowledge clicks and fifty-two percent of EBSCO clicks went to the library catalog. In general, clicks that originated in GS going to the library catalog were relatively low in comparison to EBSCO and Web of Knowledge. Only five percent of GS clicks went to interlibrary loan. Web of Knowledge and EBSCO clicks to interlibrary loan ranged from fourteen to twenty-three percent. Again, the number of clicks going to these services was relatively low by comparison. These lower numbers may indicate a preference of GS users for immediate access to online fulltext.

Table 1. 2009 Top 10 GS Targets and Ranking Comparison

Target	GS Uses	GS Rank	WOK Uses	WOK Rank	EBSCO Uses	EBSCO Rank
ScienceDirect	1674	1	342	2	872	1
Business Source	1298	2	3		24	
Academic Search	1179	3	33	9	21	
ABI/INFORM	834	4	16		696	3
Highwire Press	627	5	75	5	275	4
PsycARTICLES	250	6			1	
JSTOR Arts and Sciences I	167	7	6		265	5
Wiley-Blackwell	156	9	154	3	124	
PubMed Central	156	8	36	8	99	
Sage Publications	149	10	5		141	9

The remaining ninety-four targets were full-text resources. Google Scholar is referring our users to a vast breadth of targets. Table 1 shows the top ten targets for users coming from GS. The top resource is clearly ScienceDirect. ScienceDirect also ranks as a top resource for Web of Knowledge and EBSCO. Even as EBSCO's top target, GS users were accessing ScienceDirect almost two times more often. ScienceDirect is clearly an important resource ranking first for both GS and EBSCO.

Table 2. 2009 Google Scholar Target Category, Title Count & Usage

| Subject Area	Titles	Clicks	Titles/Clicks
Science	4,540	2,606	2
Social Science	23,917	3,091	8
Interdisciplinary	47,404	2,980	16
Humanities	4,964	80	62

All ninety-four full-text GS targets were categorized as science, social science, humanities or interdisciplinary. The interdisciplinary titles represent large packages with titles in more than one broad area. Table 2 displays each category along with the number of link resolver clicks and the number of titles in that category. All of the targets in the science category total 4,540 titles and there were 2,606 clicks from GS to a science target. In comparing the number of titles to the number of clicks, there are on average two titles for every click. The science targets are the most highly used. Although the social sciences category has the greatest number of link resolver clicks, in considering the vast number of titles in that area, use is much lower in that there is one link resolver click for every eight titles. The humanities category stands out as the least used category of targets. In this category, there were only eighty clicks to get to the 4,964 titles available. It is interesting that so much use of online journals especially in the sciences and social sciences was attributable to GS use. For our library's holdings then, GS is rather comprehensive in the coverage of these broad disciplines with the majority of holdings being in interdisciplinary collections and in the social sciences.

Interlibrary Loan Users

One way to track GS use is through interlibrary loan (ILL) requests. Interlibrary loan services are freely available to all university faculty, staff and students. As a part of the OpenURL, GS's source information is automatically transferred to the ILL form in the "cited in" field. This information is stored in the ILL transaction and can be used to identify and analyze requests

coming from GS. The increase of GS use is also apparent in ILL requests starting with only

sixty-six requests in 2006 and almost doubling every year since then. In 2009, there were 439

total requests originating from GS.

Figure 2. Google Scholar Users by Status and Discipline
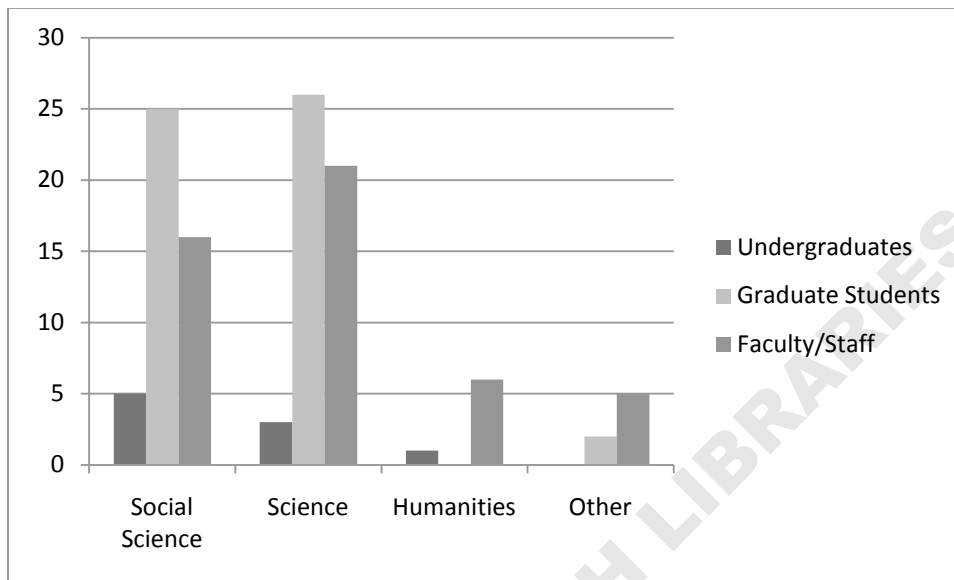
From these 439 ILL requests, there were 110 unique users. Figure 2 demonstrates the

number of users in each of the broad disciplines by user status. In looking at undergraduates, the

majority of users were in the social sciences. Although interlibrary loan services are free to all

students, undergraduates typically have the ability to switch their research topics to match the

resources readily available to them. There were more graduate student users than any other type

of users and they made up the highest number of users in both the social sciences and science

categories. Faculty were the only group represented in every discipline. Their greatest

representation was in the sciences and then the social sciences. In contrast, there were only a few

faculty and undergraduate users in the humanities.

Are there statistically significant differences between users of GS and non-GS users? In

comparing these two groups by status (N=1731), there were statistically significant differences

(*p*=.000) with more graduate and faculty/staff users than expected among GS users. In comparing the two groups by discipline, there were significantly (*p*=.001) more users in the sciences than expected and noteably fewer users in the humanities than expected. Chi-square tests related to department or combining status with discipline became problematic with cell counts less than five because of the low frequency of undergraduate GS users. In comparing the percent of ILL requests to the percent of GS ILL requests, the following departments were identified as using GS more: Business, Chemistry, Exercise Science, Mathematics, Physics, Biology, Pharmacy, Psychology and International Studies.

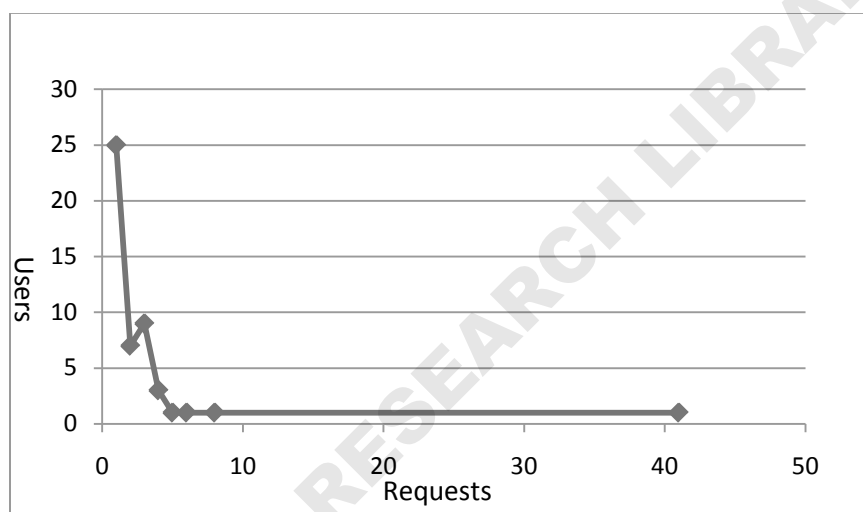Figure 3. Users by ILL Scholar Requests

Figure 3 plots the number of unique users by the number of ILL requests they placed. On the top of this curve, twenty-five users placed only one interlibrary loan request from GS. On the other end of this curve, one superuser placed over forty requests. This graph brings to mind the Pareto Principle also known as the 80/20 rule or the principle of uneven distribution where a subset of users are disproportionately productive.[xlii] This general principle of uneven or unequal distribution is the underlying premise for many laws in bibliometrics including Bradford's Law. However Bradford's Law applies primarily to serials and was discovered through bibliographic

analysis whereas the 80/20 rule's application to libraries was discovered by studying use patterns.[xliii] This first application of the 80/20 rule to libraries was done by Richard W. Trueswell in examining use of library collections.[xliv] As noted by Crawford, the Pareto Principle "holds true in an astonishingly wide variety of fields, including many aspects of librarianship." Since this study relates to use, the 80/20 rule will serve well as a means to identifying a subset of superusers. In ranking users by number of click-throughs, it was possible to determine the top twenty percent.

Applying the 80/20 rule to analyzing the 110 interlibrary loan users resulted in identifying twenty-two superusers. This was accomplished by ranking the users by their number of requests and selecting the top twenty percent. These twenty-two superusers accounted for 284 requests of the 439 requests. In other words, sixty-five percent of the GS ILL requests came from twenty percent of the users. This figure is further away from the 80/20 rule but still consistent with the laws of uneven distribution. The majority of superusers (17) were graduate students in the sciences and social sciences. Two users were undergraduates and three users were faculty. One hundred and thirty-two requests came from graduate students in the sciences. The departments/schools affiliated with these superusers included in descending order: Exercise Science, Pharmacy, Psychology, Chemistry, Physics, Biology, Engineering, Education, International Studies, Chemistry, and Business. In comparing GS superusers to non-superusers, there were statistically significant differences ($p=.008$) with more graduate students than expected and fewer Faculty/Staff and undergraduates than expected.

Another question asked is to what extent are GS users using other library databases? Seventy-six percent of ILL/GS users also used other library resources according to the "cited in" field of ILL requests. Only twenty-four percent of ILL/GS users were exclusively citing GS.

Many of these users did use Amazon, Wikipedia or some other free web resource, but these were not counted as a library resource. Citing from a known publication was also not counted as a library resource. Overall, the vast majority of ILL/GS users were using other library resources in addition to GS. To analyze these exclusive GS citers further, we looked for statistically significant differences between them and non-exclusive citers. In analyzing exclusive citers by status, significance was found (p=.008) with more exclusive citers found among graduate students.

Library Website Click-Throughs

The University of Mississippi library website uses click-throughs to provide uniform access to online databases, to assist with troubleshooting, and for usage purposes. In January 2009, we added GS to the list of library databases on the library website and established a click-through for it. The URL was proxied to allow off-campus users the ability to perform GS searches and automatically receive the library links for our institution. Google Scholar's addition to the library website was not announced or taught in library instruction.

For 2009, there were 801 click-throughs to GS from the library website. 355 were on-campus and 446 were off-campus. For the 355 on-campus click-throughs to GS, the majority (178) were coming from science buildings based on IP address. Click-throughs coming from the School of Pharmacy accounted for 147 click-throughs.

Off-campus users yielded more information since status and discipline could be determined. Several of the 446 off-campus click-throughs were excluded from analysis because they represented spiders or unauthenticated users. The remaining 390 library website click-throughs were analyzed. Many of the click-throughs were attributable to repeat GS users. Of the

390 off-campus click-throughs, there were seventy-nine unique users. With such a small sample

size, data for library website click-throughs did not lend itself to statistical testing due to cell

count issues that were low among faculty and in the humanities.

Figure 4. Off-Campus GS Library Website Users
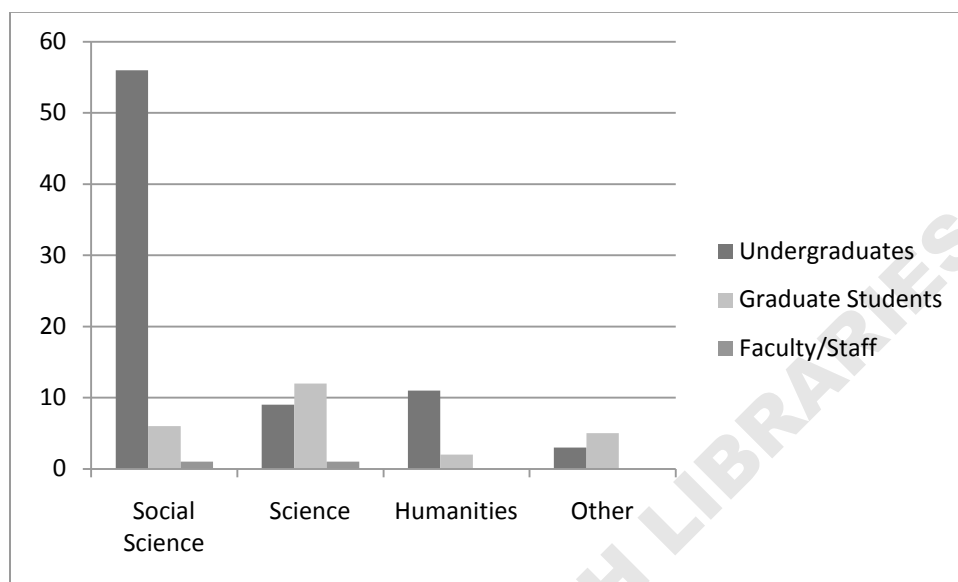
For this sample represented in Figure 4, it is interesting to note that of the seventy-nine GS users,

the majority (52) were undergraduates. Undergraduate students represent the largest user

population the library serves and are well-represented in this sample.

The majority of unique undergraduate users (56%) were coming from the social sciences

including Communicative Disorders, Education, and Psychology. In looking at graduate

students, the majority of unique users (48%) were coming from the sciences especially in the

areas of Pharmacy and Biology. In looking across the subject areas, forty-six percent of users

were in the social sciences and only twenty-eight percent were in the sciences. With only two

faculty members logged as using the GS click-through on the library website, they were clearly

underrepresented in this sample and clearly do not use this entrance to GS from off-campus.

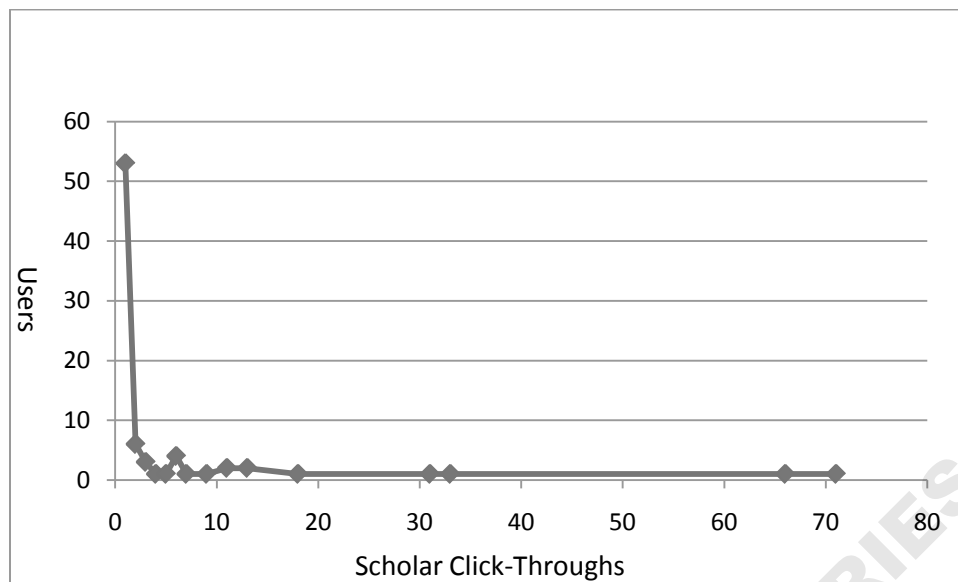Figure 5. GS Users by Library Website Click-Throughs

Figure 5 plots the number of users versus the number of GS library website click-throughs. The majority of users are at the top of the curve where fifty-three users only clicked GS one time. In the middle of the curve, there were several users clicking GS a few times. Toward the end of the curve, there were notable users or superusers who used GS more than a dozen times and up to seventy-one times.

Sessions & Resources

To facilitate further analysis, off-campus sessions were identified. Sessions were defined using the click-through log by matching patron record number with the date and time. A session was considered a continuous connection between the user and activity related to the click-through log. The default timeout period for a session was thirty minutes of inactivity. Defining a session helped analyze user behaviors within a session such as what resources were being used and when. In doing this for off-campus click-throughs, there were 348 sessions. With 390 website click-throughs, this meant several users were using GS multiple times within a session.

Forty percent of the sessions indicated the use of GS along with other library resources. For the University of Mississippi library which officially positions GS as a complementary database, this is an encouraging finding. Sixty-three of these 140 sessions concluded with the user in GS. By way of contrast, only fifteen sessions ended with the user having last clicked on Academic Search Premier. This statistic could lend itself to several interpretations. It could be that these users are performing exhaustive searches and end with GS as the broadest possible search. It is also possible that some or many of these sessions indicate that some users are not finding what they are looking for and GS could be a last resort. Others may be using GS to search for more recent information than that indexed in library subscription databases. Of the 348 sessions, 253 sessions or seventy-three percent were tied to unique users considered to be repeat users. For this reason, it is important to look at sessions in the context of unique users.

Table 3. Users & Session Pattern Category

Session Category	Users
GS after another resource	52% (41)
Multiple approaches	26% (21)
GS Exclusive	14% (11)
GS first and then other resources	8% (6)

Of all the website click-through sessions, there were seventy-nine unique users. These users were assigned session categories based on the predominant patterns of their sessions (Table 3). Fifty-two percent of GS users coming from the library website clicked GS after using another library resource. Of these GS users, thirty-two were undergraduates and twenty were in the social sciences. A smaller group of twenty-one users had a pattern of employing multiple approaches. These users were primarily in the sciences and graduate students. Because of the many GS-only sessions, it would have been easy to think that there were many users who used GS exclusively. As it turned out, there were only eleven users that used GS exclusively and they were mostly

one-time users. GS exclusive users were predominantly undergraduates in the social sciences with one exceptional superuser in the sciences. In the great minority, only six users used GS first and then moved on to other resources. They were all one-time users rather than repeat users. In general, this is evidence that users are searching traditional library resources along with GS.

Click-Through Superusers

As noted in Figure 2, a small number of users accounted for a significant amount of use. Using the 80/20 rule as a guide, it was possible to identify this subset of superusers. In ranking users by number of website click-throughs, it was possible to determine the top twenty percent. These sixteen users accounted for 264 sessions or seventy-six percent of the total sessions. In other words, the traditional 80/20 rule proved positive with seventy-six percent of use derived from twenty percent of users. These users have clearly adopted GS as a primary resource in their research although not necessarily as an exclusive resource.

In reviewing the status and disciplines associated with these superusers, there are several characteristics of this group worth discussing. Although there were more undergraduate users overall, there were only five undergraduate superusers. The top ranked superuser fell into this undergraduate group with sixty-seven sessions and came from the social sciences. There were eleven graduate superusers with the vast majority being in the sciences. The superusers ranking second to fifth were all in the sciences with sessions ranging from thirteen to sixty-three. For all of these superusers, departments associated include International Studies, Pharmacy, Biology, Engineering, Health/Exercise Science, History, Sociology/Anthropology, Classics and Education.

Google Keywords

Google Analytics tracks use of the library's link resolver web page. Users searching Google may receive GS results. This may be the entrance point to GS for many users. Google Analytics provides keywords for analysis. Keywords are defined by Google as a "word searched for using an external search engine." [xlv] Keywords coming from Google Analytics then would only be keywords where a user clicked a GS result followed by accessing the library's full text. It is not a complete listing of all search keywords for Google or Google Scholar. For 2009, 6,363 unique keywords were reported with 345 unique keywords resulting in multiple visits. Best attempts at a broad disciplinary categorization of these 345 keywords were made. Of the 345 unique keywords, the majority (54%) were in the social sciences. Thirty-eight percent of the keywords were in the sciences while only two percent were in the humanities. Six percent of the keywords did not lend themselves to a category. During the process of assigning keywords to the broad disciplines, two subject areas emerged.

The first subject area that became increasingly apparent was a broad concept of health. As an interdisciplinary subject, health could pertain to a multitude of departments and schools including Pharmacy, Exercise Science, Psychology, Biology, Nutrition & Hospitality Management, Exercise Science, Communicative Disorders, Business, Education, Engineering, Law, Political Science, and Sociology/Anthropology. This expanded concept of health is even broader than the health sciences concept. It extends to both the sciences and social sciences but includes those areas not traditionally considered a health science such as engineers developing medical equipment, the business side of hospital management, as well as legal research related to health legislation. Of the top keywords analyzed, thirty-eight percent of them related to a broad concept of health. The majority or sixty-seven percent of these health keywords were in the sciences. However, thirty-three percent of them were in the social sciences. Many of the

keywords related to health were difficult to categorize as science or social science. This interdisciplinary area could be a reason for adopting a broad-ranging tool such as GS over or in addition to more subject-specific library databases. In fact, GS could be used to help a user determine what other subject areas and related library resources to search.

The second area of subject concentration revolved around business/accounting keywords. Eighty-two of the 186 social sciences terms were in this area. Together then the broad areas of business/accounting and health made up sixty-seven percent of the social sciences keywords. Another way to look at both of these health and business concepts is that there were many current topics in those areas in 2009. These topics included health care reform, auditing standards and economic reform related to corporate buyouts and governance. New models in the areas of communication and psychological research were also evident. These might be good examples of users looking for the most up-to-date information. Other observations coming from keyword analysis were that literary authors and characters were easy to identify as well as acronyms. Additionally, there were a number of keywords that contained a name or names and date. The results of these keyword searches made it evident that GS was being used to look up a citation with minimal information.

Conclusion

GS use has been increasing at a fast pace. In comparing library link resolver use, it is second only to our many EBSCO databases and accounted for twenty-seven percent of use in 2009. In reviewing the targets GS has referred our users to, coverage for all three broad disciplines is extensive, making it a useful tool for a wide range of users. In comparing link resolver clicks to the ILL request form and the library catalog from GS and EBSCO, these services are used less often by GS users. This may indicate a preference for immediate access to online full text.

GS users identified by library website click-throughs and ILL requests represent the tip of the iceberg. For example, there were extensively more clicks coming from Google Scholar to library resources (10,522) than from users on the library website to Google Scholar (801). Although the two convenience samples were relatively small, together they help develop a picture of GS users. Both show higher use in the social sciences and sciences than in the humanities. These findings are not particularly surprising. Case's research on information searching behavior states the traditional view that the primary sources for scientists are journal articles whereas for humanists, the sources tend to be books and archives.[xlvi] In a more recent study in 2006, Gardiner concluded that humanists studying English literature used printed information more than electronic resources.[xlvii] On the other hand, the 2009 Ithaka faculty survey concluded that humanists "have been later and slower to change in many ways than their peers in the sciences, to be sure."[xlviii] In tracking changes at a disciplinary level since their 2000 survey, they have found the humanists "have demonstrated that they are on basically the same trajectory as scientists, simply less far along."[xlix] We have experienced this early adoption of electronic technologies by scientists followed by an eventual adoption among humanists at the University of Mississippi Libraries in relation to Interlibrary Loan electronic delivery.[l] It will be interesting to see in several years if the use of Google Scholar at a disciplinary level will change significantly.

The samples differ rather drastically in comparing user types although graduate students were well-represented in both. For the off-campus library website user sample, undergraduates made up the majority of users although they tended to be one-time users. The underrepresentation of faculty in this sample was rather notable and might be explained by faculty performing research on-campus from their offices rather than off-campus. They may also

have been going directly to GS instead of using the library website or relying on staff and research assistants to perform searching. For the ILL sample, graduate students and faculty were the predominant users. Taking into consideration the preference for immediate access as suggested in comparing link resolver use of ILL and library catalog services, it stands to reason that undergraduates would not be well-represented in the ILL sample.

In looking at the use of other library resources, both samples indicated that a majority of users were using other library resources along with GS. Eighty-six percent of off-campus library website users used other library resources that year and sixty-seven percent of the ILL/GS users used other library resources within a session. For the off-campus library website sessions, users also typically went to GS after using another library resource.

The laws of uneven distribution proved positive for both samples as well in that a few users were generating the vast majority of GS use. Graduate students in the sciences and social sciences were the predominant superusers. Departments that were repeatedly identified through both data sets and among superusers included Pharmacy, Biology, Engineering, Exercise Science and Psychology.

In reviewing the repeated keywords that led users from GS to the library link resolver, social sciences keywords made up the majority (54%) with science keywords making up thirty-eight percent of the total sampled. The interdisciplinary area of health was identified as a significant part of what users were searching. GS was also used for searching current topics and for citation searching. Studies of GS coverage might also be advanced by using keywords coming from Google Analytics' tracking of their link resolver to determine the adequacy of content coverage.

Science users were the top ILL/GS users. On-campus use of the library website clearly indicated GS use from science buildings. There were also more supersusers for both samples in the sciences typically among graduate students. Thirty-eight percent of keywords were attributable to science topics although it would be higher if all aspects of "health" were treated as a science topic. In looking at science titles accessed, there were fewer titles in the sciences but more use per title indicating a more concentrated use in the sciences. Several departments repeat throughout both samples: Pharmacy, Exercise Science, Communicative Disorders, Psychology and Biology. In considering these departments along with the prevalence of health-related keywords, it would seem that GS's ability to easily bridge these areas may be another reason for user adoption. Future research might include a user survey to see what behaviors users self-report regarding the use of library resources along with GS.

The results of this exploratory study could be discussion points in several areas for libraries. Internal training to help library employees stay up-to-date on Google Scholar and its features may prove helpful as a way to stay in tune with user needs. Literature related to GS instruction primarily focuses on students.[li] Based on the prevalence of graduate and faculty users, outreach to them might also be an important consideration to ensure their understanding of both the usefulness and limitations of GS in their areas.

For collection development, librarians may need to consider the "openness" of full-text resources for new purchases. The relevance of GS should also be considered during decisions related to databases and discovery tools. With many budgets in crisis, developing an understanding of the prevalence of Google Scholar use could prove helpful in striking a balance between investing in discovery tools versus licensing additional electronic content. In fact, the most recent Ithaka faculty survey indicated the increased perception of the library as a

"purchasing agent" and a decreased perception of it serving as a starting point for research.[lii]

When facing budgetary crises, Subject Librarians in the sciences may be forced to consider the usefulness of GS over less-used citation databases. The ability of link resolvers to work well with GS and keep users successful is important. GS is playing an important role as an independent discovery tool and also increasing use of library full-text resources. Although GS is free, it does reveal library-licensed full text. In this way, it helps bridge the user on the free web to library resources. It might be better to view GS as a hybrid resource rather than simply a search engine.

In conclusion, Google Scholar is successfully leading information seekers to library resources even without user instruction. The appearance of GS search results within the main Google search engine has created a pathway for Google users to use the library's resources without their realizing it. This development also gives GS a distinct advantage over library databases. Meeting users where they are is beneficial to both users and the library. The library benefits by making its licensed content more available to users out on the free web and increasing use of their licensed content. Content providers who wisely make their content accessible to search engines like GS may also be increasing use of their content, which has become increasingly important as budgets shrink and cost-per-use studies abound. Linking from Google to the library's resources is a clear benefit to our users – especially to users who perform their research outside of the library website. If the library's 2009 LibQual data was a representative sample, an estimated 9,189 undergraduates use search engines like Google daily.[liii] With so many search engine users, it makes sense to try and meet the needs of these users. In this context, GS plays an important role in bringing researches back to library-licensed full text.

Notes

[i] Institutional Research & Assessment. " Institutional Research & Assessment Data, 2008-2009," University of Mississippi. http://www.olemiss.edu/depts/university_planning/institutional_research/index.html.

[ii] Peter Jacso, "Google Scholar: the pros and the cons," *Online Information Review* 29, no. 2 (February 15, 2005): 208-214.

[iii] Karen A. Hartman and Laura Bowering Mullen, "Google Scholar and academic libraries: an update," *New Library World* 109, no. 5 (2008): 212.

[iv] Robert Schroeder, "Pointing Users Toward Citation Searching: Using Google Scholar and Web of Science," *portal: Libraries & the Academy* 7, no. 2 (April 2007): 244.

[v] Hartman and Mullen, "Google Scholar and academic libraries: an update," 212-213.

[vi] Barbara Quint, "Changes at Google Scholar: A Conversation with Anurag Acharya," *Journal of Library Administration* 47, no. 1 (2008): 77.

[vii] Hartman and Mullen, "Google Scholar and academic libraries: an update," 213.

[viii] Chris Neuhaus, Ellen Neuhaus, and Alan Asher, "Google Scholar Goes to School: The Presence of Google Scholar on College and University Web Sites," *The Journal of Academic Librarianship* 34, no. 1 (January 2008): 41.

[ix] Alireza Noruzi, "Google Scholar: The New Generation of Citation Indexes," *Libri: International Journal of Libraries & Information Services* 55, no. 4 (December 2005): 174.

[x] Chris Neuhaus et al., "The Depth and Breadth of Google Scholar: An Empirical Study," *portal: Libraries & the Academy* 6, no. 2 (April 2006): 127-141.

[xi] Quint, "Changes at Google Scholar: A Conversation with Anurag Acharya," 78.

[xii] Anne-Wil Harzing and Ron van der Wal, "A Google Scholar h-index for journals: An alternative metric to measure journal impact in economics and business." *Journal of the American Society for Information Science & Technology* 60, no. 1 (January 2009): 41-46.

[xiii] Lokman I. Meho and Kiduk Yang, "Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar," *Journal of the American Society for Information Science & Technology* 58, no. 13 (November 2007): 2105-2125.

[xiv] Hartman and Mullen, "Google Scholar and academic libraries: an update," 215.

[xv] Neuhaus et al., "The Depth and Breadth of Google Scholar: An Empirical Study," 127-141

xvi Burton Callicott and Debbie Vaughn, "Google Scholar vs. Library Scholar: Testing the Performance of Schoogle," *Internet Reference Services Quarterly* 10, no. 3 (July 2005): 71-88.

xvii Schroeder, "Pointing Users Toward Citation Searching: Using Google Scholar and Web of Science," 243-244.

xviii Marilyn Christianson, "Ecology Articles in Google Scholar: Levels of Access to Articles in Core Journals," *Issues in Science & Technology Librarianship*, no. 50 (Winter 2007): 3.

xix Michael Levine-Clark and Joseph Kraus, "Finding Chemistry Information Using Google Scholar," *Science & Technology Libraries* 27, no. 4 (2007): 3-17.

xx John J. Meier and Thomas W. Conkling, "Google Scholar's Coverage of the Engineering Literature: An Empirical Study," *Journal of Academic Librarianship* 34, no. 3 (May 2008): 201.

xxi William H. Walters, "Google Scholar Search Performance: Comparative Recall and Precision," *portal: Libraries and the Academy* 9, no. 1 (January 2009.): 5.

xxii Jared L. Howland et al., "How Scholarly Is Google Scholar? A Comparison to Library Databases," *College & Research Libraries* 70, no. 3 (May 2009): 227-234.

xxiii Schroeder, "Pointing Users Toward Citation Searching: Using Google Scholar and Web of Science,"

xxiv S. Gardner and S. Eng, "Gaga over Google? Scholar in the Social Sciences," *Library Hi Tech News* 22, no. 8 (January 1, 2005): 42-45.

xxv Howland et al., "How Scholarly Is Google Scholar? A Comparison to Library Databases," 227-234.

xxvi Meho and Yang, "Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar," 2105-2125.

xxvii Walters, "Google Scholar Search Performance: Comparative Recall and Precision," 218.

xxviii Neuhaus, Neuhaus, and Asher, "Google Scholar Goes to School: The Presence of Google Scholar on College and University Web Sites," 49.

xxix Glenn Haya, Else Nygren, and Wilhelm Widmark, "Metalib and Google Scholar: a user study," *Online Information Review* 31, no. 3 (June 26, 2007): 365.

xxx Xiaotian Chen, "MetaLib, WebFeat, and Google: the strengths and weaknesses of federated search engines compared with Google," *Online Information Review* 30, no. 4 (January 1, 2006): 413.

xxxi Gail Herrera, "Meta Searching and Beyond: Implementation Experiences and Advice from an Academic Library," *Information Technology & Libraries* 26, no. 2 (June 2007): 51.

xxxii Jillian R. Griffiths and Peter Brophy, "Student Searching Behavior and the Web: Use of Academic Resources and Google.," *Library Trends* 53, no. 4 (Spring2005 2005): 545.

xxxiii Cathy De Rosa et al., "College Students' Perceptions: Libraries & Information Resources [OCLC - Reports]," 2005, 6-2, http://www.oclc.org/reports/pdfs/studentperceptions.pdf.

xxxiv Ibid., 6-3.

xxxv Lotta Haglund and Per Olsson, *The Impact on University Libraries of Changes in Information Behavior Among Academic Researchers: A Multiple Case Study,* 34 52-59, 52 (2008).

xxxvi Ibid., 34:55.

xxxvii Ibid.

xxxviii Roger C. Schonfeld and Ross Housewright, "Faculty Survey 2009: Key Strategic Insights for Libraries, Publishers, and Societies" (Ithaka S+R, April 7, 2010), 7, http://www.ithaka.org/ithaka-s-r/research/faculty-surveys-2000-2009/faculty-survey-2009.

xxxix Ibid., 24.

xl Xi Niu et al., "National study of information seeking behavior of academic researchers in the United States," *Journal of the American Society for Information Science & Technology* 61, no. 5 (May 2010): 879.

xli Schonfeld and Housewright, "Faculty Survey 2009: Key Strategic Insights for Libraries, Publishers, and Societies," 4.

xlii Thomas E. Nisonger, "The "80/20 Rule" and Core Journals," *Serials Librarian* 55, no. 1 (2008): 78.

xliii Ibid., 74.

xliv Richard L. Trueswell, "Some Behavioral Patterns of Library Users: The 80/20 Rule," *Wilson Library Bulletin* 43, no. 5 (January 1969): 458-461.

xlv Schonfeld and Housewright, "Faculty Survey 2009: Key Strategic Insights for Libraries, Publishers, and Societies," 4.

xlvi Donald O. Case, *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior* (Amsterdam: Academic Press, 2002).

xlvii Donna Gardiner, David McMenemy, and Gobinda Chowdhury, "A snapshot of information use patterns of academics in British universities," *Online Information Review* 30, no. 4 (July 2006): 341-359.

xlviii Schonfeld and Housewright, "Faculty Survey 2009: Key Strategic Insights for Libraries, Publishers, and Societies," 34.

xlix Ibid.

l Gail Herrera, "Interlibrary Loan User Behaviors in an Academic Library," *Journal of Interlibrary Loan, Document Delivery & Information Supply* 14, no. 2 (December 2003): 19-36.

li William Badke, "Google Scholar and the Researcher," *Online* (June 2009).

lii Schonfeld and Housewright, "Faculty Survey 2009: Key Strategic Insights for Libraries, Publishers, and Societies," 9.

liii "LibQual+ 2009 Survey. University of Mississippi, Oxford" (Association of Research Libraries, Texas A&M University, 2009).

Appendix A. EBSCO Databases that are Link Resolver Sources

Academic Search Premier, Agricola, Alt HealthWatch, America: History & Life, Art & Architecture Complete, Biological Abstracts, Business Source Complete, Communication & Mass Media Complete, Computer Science Index, Consumer Health Complete, Environment Complete, ERIC, GeoRef, GreenFILE, Historical Abstracts, Hospitality & Tourism Complete, Humanities International Complete, Inspec, Legal Collection, Library, Information Science & Technology Abstracts, Literary Reference Center, MEDLINE, MLA Directory of Periodicals, MLA International Bibliography, Newspaper Source, Philosopher's Index, Professional Development Collection, Psychology and Behavioral Sciences Collection, PsycINFO, Religion and Philosophy Collection, RILM Abstracts of Music Literature, Shock & Vibration Digest, SocINDEX with Full Text, SPORTDiscus with Full Text