



Faculty Publications

2008-06-09

How Scholarly Is Google Scholar? A Comparison of Google Scholar to Library Databases

Jared L. Howland
jared_howland@byu.edu

Thomas C. Wright
tom_wright@byu.edu

Rebecca A. Boughan
rebecca_boughan@byu.edu

Brian C. Roberts

Follow this and additional works at: <https://scholarsarchive.byu.edu/facpub>



Part of the [Library and Information Science Commons](#)

Original Publication Citation

null

BYU ScholarsArchive Citation

Howland, Jared L.; Wright, Thomas C.; Boughan, Rebecca A.; and Roberts, Brian C., "How Scholarly Is Google Scholar? A Comparison of Google Scholar to Library Databases" (2008). *Faculty Publications*. 182.
<https://scholarsarchive.byu.edu/facpub/182>

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.

Brigham Young University

Harold B. Lee Library

Institutional Repository

<https://dspace.lib.byu.edu>

How Scholarly Is Google Scholar? A Comparison of Google Scholar to
Library Databases

Howland, Jared L; Wright, Thomas C; Boughan, Rebecca A; Roberts, Brian C
Harold B. Lee Library

Submitted to Repository on: 2008-06-09T20:19:03Z

Please use this identifier to cite or link to this item: <https://dspace.byu.edu/handle/1877/634>

[Copyright Information](#)

<https://dspace.byu.edu/handle/1877/634>

How Scholarly Is Google Scholar?

A Comparison of Google Scholar to Library Databases

Jared L. Howland, Thomas C. Wright, Rebecca A. Boughan, and Brian C. Roberts
May 21, 2008

Abstract

Google Scholar (GS) was released as a beta product in November of 2004. Since then, GS has been scrutinized and questioned by many in academia and the library field. Our objectives in undertaking this study were to determine how scholarly GS is in comparison with traditional library resources and to determine if the scholarliness of materials found in GS varies across disciplines. We found that GS is, on average, 17.6% more scholarly than materials found only in library databases and that there is no statistically significant difference between the scholarliness of materials found in GS across disciplines.

Introduction

Google Scholar (GS) was introduced to the world in November of 2004 as a beta product. It has been embraced by students, scholars, and librarians alike. However, GS has received criticism regarding the breadth and scope of available content. We undertook this study to answer two questions regarding these common criticisms: (1) Are GS result sets more or less scholarly than licensed library database result sets? and (2) Does the scholarliness of GS vary across disciplines?

Literature Review

GS, which is still branded as a beta version, has not only become a common fixture in library literature but is also becoming ubiquitous in information-seeking behavior of users. GS was initially met with curiosity and skepticism [1-3]. This was followed by a period of systematic study [4-7]. More recently, there has been optimism about GS's potential to move us towards Kilgour's goal of 100% availability of information [8]. Librarians find themselves reluctantly acknowledging users' preferences for one-stop information shopping by giving GS ever-increasing visibility on their web pages [9]. Even as they begrudgingly promote GS, the debate continues within the information community as to the advisability of guiding users to this tool. The view of critics like Péter Jascó [10-12], who use terms such as "shallowness" and "artificial unintelligence" to describe the program, seems to be giving way to a landscape where respected publishers (e.g., Cambridge) and platforms (e.g., JSTOR) are now offering links out to GS for more citations.

Early studies of GS tried to match citations "hit to hit" in comparison with traditional academic search engines. Jascó even provided a web site where the curious could compare search results between GS and the likes of Nature, Wiley or Blackwell [13]. More recently, studies have appeared that track the "value-added" open access citations that appear uniquely in

How Scholarly Is Google Scholar?: A Comparison of Google Scholar to Library Databases

Jared L. Howland, Thomas C. Wright, Rebecca A. Boughan and Brian C. Roberts

GS versus other sources [14]. However, is comprehensiveness of content the primary indicator of a resource's usefulness?

Every title from every database may not be in GS, but that should not be an indictment of GS's inability to return scholarly results across disciplines. The algorithms GS uses to return result sets cannot really be compared to library database algorithms. However, what is returned can be judged for its relevancy and "scholarliness." Up to this point, studies of GS have followed the example of Neuhaus et al. (2006), which compared GS content to forty-seven other databases. This title-by-title and citation-by-citation comparison is a pure numerical measure but neglects to address the efficacy of any particular search or the scholarly nature of content or algorithms in discovering that content.

We felt that a different approach was needed. Rather than measuring what has gone into the database, we have sought, to some degree, to evaluate what comes out as a result of search queries. We have done this by involving subject librarians with knowledge of typical reference questions and using those questions to query both GS and discipline-specific databases. We then asked the same librarians to judge the search results using a rubric of scholarliness. This notion of scholarliness utilizes a common collection-assessment tool as outlined by Alexander [15]. This model considers many factors, including accuracy, authority, objectivity, currency, and coverage. In this way we have attempted to inject a qualitative value of GS results to the ongoing debate.

Methodology

We selected seven subject librarians from various academic disciplines, humanities, science and social science. Each specialist was blind to the purpose of the study. We requested that they provide us (1) a sample question that they typically receive from students, (2) a structured query to search a library database, and (3) the library database they would use for that particular query. [INSERT TABLE 1]

We then used their data in two different ways. First, we translated the library database query into an equivalent search string used by GS. Using the original query and the query translated to work with GS, we searched both the library database and GS and retrieved the citations and full text for the first thirty results. We selected thirty results because research has shown that less than one percent of all users ever go to a third page of results and most search engines return about ten results per page [16].

Next, we took the citations from the library databases and determined if they could also be found using GS and took the citations from GS to see if they could also be found in the library database. This allowed us to calculate the overlap of citations between the library databases and GS.

We standardized the formatting of the citations and inserted them randomly into a spreadsheet, which contained a rubric that was used to assign a scholarliness score to each of the citations. The rubric contained six criteria, based on Alexander's (1999) model of evaluating resources, to judge scholarliness: (1) accuracy, (2) authority, (3) objectivity, (4) currency, (5) coverage, and (6) relevancy. These criteria were graded on a scale of 1 (below average) to 3 (above average) and summed to create a total scholarliness score for each citation. [INSERT TABLE 2]

We provided the subject librarians with the full text of each of the citations and asked them to use the rubric to evaluate the scholarliness of the individual citations. After the grading was completed, we were able to group each citation from the subject librarian into one of three

How Scholarly is Google Scholar?: A Comparison of Google Scholar to Library Databases

Jared L. Howland, Thomas C. Wright, Rebecca A. Boughan and Brian C. Roberts

categories: (1) the citation was available only in the library database, (2) the citation was available only in GS, or (3) the citation was available in both the library database and GS. We have used the term “exclusivity” to describe the three categories.

Once we had grouped the citations by category, we ran a statistical analysis that controlled for the effect of the individual librarian on the total scholarliness score, for the effect of “exclusivity”, and for any interaction there may have been between both librarian and “exclusivity”:

$$\text{total scholarliness score} = \mu + E_i + L_j + EL_{ij} + \epsilon_{ijkl}$$

Where:

μ	=	Average total score
E	=	Effect due to “exclusivity” ($i = 1, 2, 3$)
L	=	Effect due to librarian ($j = 1, 2, \dots, 7$)
EL	=	Interaction between “exclusivity” and librarian
ϵ	=	Error term

Within the context of this formula, E controls for any effect due to the “exclusivity” of the citation, where i represents each of the three categories of “exclusivity” (i.e., the citation was found only in the database, it was found only in GS, or it was found in both the database and GS). Each librarian (L) also played a role in the total scholarliness score. One librarian could have provided consistently low scores with another having a tendency toward higher scores. To account for this disparity, each librarian was treated as a factor in the total scholarliness score, where j represents each of the seven participants. In short, this formula allowed us to calculate a measure of scholarliness while accounting for differences in where the citations were located and between librarians.

Results

The mean scholarliness score of citations found only in GS was 17.6% higher than the score for citations found only in licensed library databases. In fact, across all but one of the tested disciplines, citations found only in GS had a higher average scholarliness score than citations found only in licensed library databases. The one discipline with a lower score, however, had only two unique citations in the library database, so the exact significance of the scores for that discipline is imprecise. Additionally, the citations found in both GS and licensed library databases had a higher average score than citations found only in one or the other [INSERT TABLE 3]. Finally, there was no statistically significant difference found between the scholarliness score across disciplines within GS. Searching for either a humanities topic or a science topic yielded no difference in the scholarliness score of citations discovered in GS.

Discussion of Results

It is interesting to note that there was very little overlap between the initial thirty citations returned by the databases and the initial thirty citations returned by GS. In fact, only one query of the seven had any overlapping citations between GS and the database— an overlap of five citations from JSTOR that appeared within the first thirty results in GS. Despite this initial lack of overlap, once we began to search for specific citations, we found that GS actually contained 76% of all the citations found in the library databases, while the library databases contained only 47% of the citations found in GS. [INSERT TABLE 4]

This seems to validate the decision of many students to use Google first to look for information. If GS contains much of the content available in library databases, why shouldn't

How Scholarly Is Google Scholar?: A Comparison of Google Scholar to Library Databases

Jared L. Howland, Thomas C. Wright, Rebecca A. Boughan and Brian C. Roberts

students begin where the most content exists? The argument is made that GS will return millions of hits, many of which are spurious at best, while a library database will only return a few thousand results that are more focused to the query.

However, the power of ordering results by relevancy, combined with the fact that very few people ever go beyond the first page of results, creates a searcher-imposed higher level of precision for any search engine. This is particularly true of GS, where the most relevant and more scholarly, material floats to the top of the list, while the less precise material falls to the bottom, where it is rarely seen. Hit counts are of secondary importance in a GS search; the key to GS's success is relevancy ranking and a large universe of information.

A database is limited to its defined title list of content, whereas GS, by its very nature, is open to a much broader set of content that aids the researcher. Business Source Premier, one of the library databases, was the only library database where we found more GS citations in the database than database citations in GS. However, even in this one instance, the scholarly score for citations found only in GS was higher than the score for citations found only in the database. The citations found in both GS and the databases received even higher scores and these citations were only exposed through the first 30 hits in GS. This seems to indicate that even when GS is returning fewer titles, as in this case with Business Source Premier, it still returns citations that are more scholarly to the top.

Up to this point, many of our library databases have defaulted to sorting by date rather than by relevancy. The fact that many databases are now adding relevancy search options seems to indicate that GS got it right in the first place. It appears that GS has done a better job of both precision and recall than library databases have.

Many studies have compared content in library databases to content in GS and found inconsistencies. The purpose of both search systems, however, is to discover relevant, scholarly content. Using our scholarliness model, we found that, across disciplines, GS is generally superior to individual databases in retrieving appropriate citations. As more publishers share their content with GS, we would expect the effectiveness of a GS search to increase.

Future Studies

The statistical results from this study can be extrapolated only to the specific topics and subject librarians that were involved in the study. A more comprehensive statistical methodology would need to be constructed in order to make the results generally applicable. However, our results were compelling enough to make us believe that the results would hold up to more strenuous tests. Additionally, the rubric we used in our study was only a three-point Likert scale. Finding statistically significant differences would have been easier had we selected a seven or more point Likert scale.

Additionally, our analysis used a vetted approach to evaluating scholarliness of resources. A more objective view of scholarliness could be obtained by using some variation of citation analysis (citation counts, ISI impact factor, etc.). We started to do such an analysis but decided there were too many trade-offs to be appropriate given the methodology we used for this study. For example, citation counts are difficult to come by for materials other than journal articles, and impact factors are calculated for journals only and not for specific articles. Alternate methodologies would likely be able to overcome or account for the shortcomings of using citation analysis to judge scholarliness.

Our study used skilled librarians to create search queries and to judge the quality of the citations retrieved. Unlike most students, the librarians used complex search queries to find more

How Scholarly is Google Scholar?: A Comparison of Google Scholar to Library Databases

Jared L. Howland, Thomas C. Wright, Rebecca A. Boughan and Brian C. Roberts

relevant results. Students would be more likely to use natural language queries to find citations. Complex search queries could return very different results from natural language queries. Future studies will need to address the potential differences to find out if the results we found hold across different types of searches.

Finally, future studies need to look at the appropriateness of comparing GS to individual library databases. It is probable that federated searching is more comparable to GS than are individual library databases. However, how users and librarians select which resources to use in a federated search and how the federated search engine returns the results would still impact the discoverability of scholarly resources. Some studies have already started down this road [17], but GS result sets have still not been carefully compared to result sets from federated search products.

Libraries have begun to build local GS's, using tools such as *Primo* (Ex Libris), *AquaBrowser* (Medialab Solutions), and *Encore* (Innovative Interfaces), that have the potential to aid users in discovering even more scholarly materials than what is currently discovered in GS. Comparing GS to a future system that has completely indexed all local content and content available to libraries but provided by third parties would be the ultimate comparison.

Conclusion

Typical arguments against GS focus on citation counts and point to inconsistent coverage between disciplines. We felt the more appropriate analysis was to compare the scholarliness of resources discovered using GS with resources found in library databases. This analysis showed that GS yielded more scholarly content than library databases, with no statistically significant difference in scholarliness across disciplines. Despite these findings, GS is not in competition with library databases. In truth, without the cooperation of database vendors and publishers, GS would not exist as it does today. GS is simply a discovery tool for finding scholarly information while databases still perform the function of providing access to the content unearthed by a GS search.

Bibliography

1. Brophy J, Bawden D. Is Google enough? Comparison of an Internet Search Engine with Academic Library Resources. *Perspectives* **2005**;57:498-512.
2. Gardner S, Eng S. Gaga over Google? Scholar in the Social Sciences. *Library Hi Tech News* **2005**;22:42-5.
3. Kesselman M, Watstein SB. Google Scholar and Libraries: Point/Counterpoint. *Reference Services Review* **2005**;33:380-7.
4. Bakkalbasi N, Bauer K, Glover J, Wang L. Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomed Digit Libr* **2006**;3:7.
5. Kousha K, Thelwall M. Google Scholar and Google Web/URL Citations: A Multi-Discipline Exploratory Analysis. *Journal of the American Society for Information Science and Technology* **2007**;58:1055-65.
6. Robinson ML, Wusteman J. Putting Google Scholar to the Test: A Preliminary Study. *Program: Electronic Library & Information Systems* **2007**;41:71-80.
7. Neuhaus C, Neuhaus E, Asher A, Wrede C. The Depth and Breadth of Google Scholar: An Empirical Study. *Portal: Libraries and the Academy* **2006**;6:127-41.
8. Pomerantz J. Google Scholar and 100 Percent Availability of Information. *Information Technology and Libraries*, June **2006**:52-6.
9. Mullen LB, Hartman KA. Google Scholar and the Library Web Site: The Early Response by ARL Libraries. *College and Research Libraries* **2006**;67:106-22.

How Scholarly is Google Scholar?: A Comparison of Google Scholar to Library Databases
Jared L. Howland, Thomas C. Wright, Rebecca A. Boughan and Brian C. Roberts

10. Jascó P. As We May Search - Comparison of Major Features of the Web of Science, Scopus, and Google Scholar Citation-based and Citation-enhanced Databases. *Curr Sci* **2005**;89:1537-47.
11. Jascó P. Google Scholar: The Pros and the Cons. *Online Information Review* **2005**;29:208-14.
12. Jascó P. Deflated, Inflated and Phantom Citation Counts. *Online Information Review* **2006**;30:297-309.
13. Péter Jascó: side-by-side2 - Google Scholar vs Native Search. Available at:
<http://www2.hawaii.edu/~jacso/scholarly/side-by-side2.htm>. Accessed March 24, 2008.
14. Kousha K, Thelwall M. How is Science Cited on the Web? A Classification of Google Unique Web Citations. *J Am Soc Inf Sci Technol* **2007**;58:1631-44.
15. Alexander JE. *Web Wisdom: How to Evaluate and Create Information Quality on the Web*. Mahwah N.J.: Lawrence Erlbaum Associates Publishers; **1999**.
16. Nielsen J, Loranger H. *Prioritizing Web Usability*. Berkeley, Calif.: New Riders; **2006**.
17. Chen, X. MetaLib, WebFeat, and Google: The strengths and weaknesses of federated search engines compared with Google. *Online Information Review* **2006**;30:413-27.

How Scholarly is Google Scholar?: A Comparison of Google Scholar to Library Databases

Jared L. Howland, Thomas C. Wright, Rebecca A. Boughan and Brian C. Roberts

Tables

Table 1: Academic representation in this study

Academic Discipline	Database Query	GS Query	Library Database
Science	(ACL or “anterior cruciate ligament*”) and injur* and (athlet* or sport or sports) and (therap* or treat* or rehab*)	ACL OR “anterior cruciate ~ligament” ~injury ~athlete OR sport ~therapy OR ~treatment OR ~rehabilitation	SportDiscus
Science	lung cancer and (etiolo* or caus*) and (cigarette* or smok* or nicotine*)	lung cancer ~etiology OR ~cause ~cigarette OR ~smoking OR ~nicotine	Medline
Science	“dark matter” and evidence	“dark matter” evidence	Applied Science and Technology Abstracts
Social Science	(“fast food” or mcdonald’s or wendy’s or “burger king” or restaurant) and franchis* and (knowledge n3 transfer or “knowledge management” or transfer” OR “knowledge train*)	OR wendy’s OR “burger king” OR restaurant ~franchise “knowledge management” OR ~train	Business Source Premier

How Scholarly is Google Scholar?: A Comparison of Google Scholar to Library Databases
Jared L. Howland, Thomas C. Wright, Rebecca A. Boughan and Brian C. Roberts

Social Science (“standardized test*” or “high“standardized ~test” OR PsycINFO
 stakes test*”) and (“learning “high stakes ~test” “learning
 disabilit*” or Dyslexia or ~disability” OR dyslexia OR
 “learning problem”) and “learning problem”
 accommodat* ~accomodation

Humanities (bilingual* or L2) and (child*~bilingual OR L2 ~child OR Linguistics and
 or toddler) and “cognitive toddler “cognitive Language
 development” development” Behavior
 Abstracts

Humanities (memor* or remembrance or ~memor OR remembrance JSTOR
 memoir*) and (holocaust) OR ~memoir holocaust
 and (Spiegelman or Maus) Spiegelman OR Maus

How Scholarly is Google Scholar?: A Comparison of Google Scholar to Library Databases
Jared L. Howland, Thomas C. Wright, Rebecca A. Boughan and Brian C. Roberts

Table 2: Rubric for grading scholarliness

1 = Below Average Quality; 2 = Average Quality; 3 = Above Average Quality

	Citation References	Accuracy	Authority	Objectivity	Currency	Coverage	Relevancy
Number							
1	Barnes, J. E., & Hernquist, L. E. (1993). Computer models of colliding galaxies. <i>Physics Today</i> , 46, 54–61.	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3
2	Bergstrom, L. (2000). Nonbaryonic dark matter: Observational evidence and detection methods. <i>Reports on Progress in Physics</i> , 63(5), 793–841.	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3

How Scholarly is Google Scholar?: A Comparison of Google Scholar to Library Databases
Jared L. Howland, Thomas C. Wright, Rebecca A. Boughan and Brian C. Roberts

Table 3: Scholarliness based on “exclusivity” (maximum scholarliness score is 18)

Participant	Found Only in Database Average Score	Found Only in GS Average Score	Percent Change in Scholarliness Score	Found in Both Scholarliness Score
1	11.7	16.1	36.8%	13.5
2	13.2	13.8	4.5%	14.6
3	N/A	12.0	N/A	15.6
4	10.0	13.5	35.0%	14.3
5	10.0	11.6	16.0%	11.5
6	11.7	12.8	8.5%	14.3
7	16.5	14.4	-12.7%	13.9
Least Squares Mean	11.9	14.0	17.6%	14.2

How Scholarly is Google Scholar?: A Comparison of Google Scholar to Library Databases
Jared L. Howland, Thomas C. Wright, Rebecca A. Boughan and Brian C. Roberts

Table 4: Overlap of citations

	Participant Percent of database	Percent of GS
	citations in GS	citations in database
1	76.7%	0.0%
2	83.3%	43.3%
3	100.0%	96.7%
4	96.7%	80.0%
5	93.3%	28.0%
6	0.0%	46.7%
7	81.8%	34.5%
AVERAGE	76.0%	47.0%