Google Scholar Search Performance: Comparative Recall and Precision

William H. Walters

This document is the final, published version of an article in *Portal: Libraries and the Academy*, vol. 9, no. 1 (January 2009), pp. 5–24.

It is also available from the publisher's web site at http://muse.jhu.edu/login?auth=0&type=summary&url=/journals/portal_libraries_and_the_academy/v009/9.1.walters.html

Google Scholar Search Performance: Comparative Recall and Precision

William H. Walters

abstract: This paper presents a comparative evaluation of Google Scholar and 11 other bibliographic databases (Academic Search Elite, AgeLine, ArticleFirst, EconLit, GEOBASE, MEDLINE, PAIS International, POPLINE, Social Sciences Abstracts, Social Sciences Citation Index, and SocINDEX), focusing on search performance within the multidisciplinary field of later-life migration. The results of simple keyword searches are evaluated with reference to a set of 155 relevant articles identified in advance. In terms of both recall and precision, Google Scholar performs better than most of the subscription databases. This finding, based on a rigorous evaluation procedure, is contrary to the impressions of many early reviewers. The paper concludes with a discussion of a new approach to document relevance in educational settings—an approach that accounts for the instructors' goals as well as the students' assessments of relevance.

Google Scholar (GS) has attracted substantial attention due to its potential as a free, multidisciplinary bibliographic database. Unlike most of the databases offered through libraries and other information agencies, GS does not require a subscription, registration, or payment. Because it is based on the popular Google search engine, Google Scholar has been perceived by some as a threat to library-based information services.[1] Fewer than 30 percent of North American research libraries include GS in their online resource lists, and only 5 percent include it in their public access catalogs.[2]

Published reviews of Google Scholar have tended to focus on its idiosyncrasies and shortcomings. Several authors have noted the apparent deficiencies of the GS search mechanism: the lack of controlled vocabulary for subject terms; the lack of authority control for author names and journal titles; inconsistent handling of Boolean operators; the inability to sort retrieved records by any criterion other than estimated relevance; and the absence of mechanisms for marking, manipulating, and exporting search results.[3] However, some recent studies suggest that GS performs reasonably well despite these

portal: Libraries and the Academy, Vol. 9, No. 1 (2009), pp. 5–24.

deficiencies. Relying on informal standards of relevance, Burton Callicott and Debbie Vaughn found that "Google Scholar's results in the humanities were surprisingly solid" with respect to five topics likely to be of interest to undergraduate students.[4] D. Yvonne Jones adopted a comparative approach, evaluating the performance of 10 bibliographic databases in retrieving papers on Nodilittorina, a type of periwinkle.[5] In her analysis, GS performed better than all but BIOSIS, returning more results than ArticleFirst, BasicBIOSIS, Electronic Collections Online, HighWire, MEDLINE, ProQuest, SciFinder Scholar, and WilsonWeb. Jones assumed that every search result was relevant, however, and did not examine the quality of the records retrieved by each database.

Susan Gardner and Susanna Eng compared Google Scholar with ERIC, PsycINFO, and SSCI, reporting that GS retrieved more results than any other database but that it failed to cover the most recent literature. They concluded, "There is more variety in Google Scholar and a higher number of results, but they are not necessarily as scholarly or relevant."[6] Their method of assessing relevance was based solely on the appearance of the search term (*home schooling*) in the title, abstract, or text of each article.

Evaluating the works cited in students' research papers, Rena Helms-Park, Pavlina Radia, and Paul Stapleton found that the information sources identified through Google Scholar were no different in quality than those identified through traditional bibliographic databases.[7] Specifically, the works found in GS were identical to the others on each of the four standards used in a blind assessment procedure—authority, objectivity, rigor, and transparency.

Another recent study compared GS with seven subscription databases, reporting that Google Scholar provides the most comprehensive coverage of the later-life migration literature.[8] That analysis was based on a series of title searches and examined the content of the GS database rather than the effectiveness of its search mechanism. In contrast, this paper evaluates the performance of GS and 11 other databases in retrieving relevant articles through subject keyword searches.

Specialized subject searches often make use of Boolean logic, controlled vocabulary, or other search features not available through the GS interface. However, this analysis is intended to represent the behavior of a less experienced searcher with an interest in obtaining adequate rather than comprehensive results without expending a great deal of effort. Arguably, this is the kind of searcher most likely to be familiar with the Google interface, to choose Google Scholar rather than another research-oriented database, and to assume that bibliographic search mechanisms will respond well to simple but intuitively reasonable search strategies.

Methods

Google Scholar and 11 other databases were evaluated in terms of both recall and precision. *Recall* represents one aspect of search performance—the effectiveness of each database in retrieving relevant documents. Specifically, it is calculated as the number of relevant items retrieved as a proportion of all the relevant items that might potentially be retrieved. In contrast, *precision* accounts for both the retrieval of relevant documents and the exclusion of non-relevant documents. It is calculated as the number of relevant items retrieved as a proportion of all items retrieved.

Each database was evaluated with reference to a set of 155 relevant documents identified in advance—the most important journal articles on later-life migration published from 1990 to 2000. Only those 155 papers were regarded as relevant. Potentially relevant documents were identified through database searching, citation tracing, journal browsing, and consultation with colleagues in the social sciences. More than 500 papers were considered for inclusion in the list of relevant works, but only 155 met the required standard in all five areas of assessment: subject matter, importance of findings, innovativeness of methods or approach, number of other studies published on the topic, accessibility of content (readability), and accessibility of the document itself (availability to students and scholars).[9]

Later-life migration includes elderly migration, retirement migration, post-retirement migration, and related types of geographic mobility. It was chosen as a search topic due to its multidisciplinary scope, its coverage in several major social science databases, and its appropriateness as an undergraduate term paper topic. Strictly speaking, the results of this analysis apply just to the literature of later-life migration. Nonetheless, this subject may be broadly representative of undergraduates' research topics due to its scope, its policy relevance, and its accessibility to non-specialist readers.

Each database was evaluated through a simple keyword search. Potential search terms were generated through a count of the words that appeared most often in the titles of the 155 relevant articles. *Migration* appeared 80 times, followed by *elderly* (48 times), *retirement* (21 times), *population* (20 times), and *states* (18 times). Although *later-life migration* is the most inclusive term, encompassing both elderly migration (based on age) and retirement migration (based on career status), that phrase has not been used extensively in the literature. *Elderly migration* was chosen as the search term, since it returned 1.9 times as many hits as *retirement migration* across the set of 12 databases and at least 1.5 times as many hits as *retirement migration* in every database except PAIS.

Keyword searches for *elderly migration* were conducted in each of the 12 databases: Google Scholar, Academic Search Elite, AgeLine, ArticleFirst, EconLit, GEOBASE, MEDLINE, PAIS International, POPLINE, Social Sciences Abstracts, Social Sciences Citation Index (SSCI), and SocINDEX. The search results were then used to generate recall and precision statistics for each database. Each search was undertaken using the simplest search interface that permitted the appropriate date restriction (1990 to 2000). The results were sorted by relevance, if possible, and by date otherwise. (Relevance sorting was available in nine of the 12 databases—all but AgeLine, POPLINE, and Social Sciences Abstracts.)

The appendix shows the details of the search procedures. Although the searches were designed to be as similar as possible within each database, a few significant variations can be noted. First, AgeLine uses *elderly* as a stopword and truncates each word after the seventh letter, so the phrase *elderly migration* returns all records with the character string "migrati." Second, ArticleFirst records do not include abstracts, so keyword searches in that database search only the titles, subject headings, and notes. Finally, keyword searches in Google Scholar search the full text when it is available for indexing, as well as the bibliographic records and abstracts.[10] Of the 144 relevant records included in the GS database, 25 percent have links to searchable full text. Moreover, the presence of searchable full text nearly doubles the chance that a particular GS record will

be retrieved by a keyword search for *elderly migration*. (Sixty-four percent of relevant GS records with full text are retrieved by that search, compared to just 34 percent of relevant GS records without full text.)

This analysis, based on a set of 155 papers published before 2001, does not examine Google Scholar's effectiveness in retrieving recently published items. Although several reviewers have criticized GS for its infrequent update schedule,[11] this study does not evaluate the availability of recent items. Likewise, it does not assess Google Scholar's effectiveness as a citation-tracking database.[12]

General Findings

Google Scholar generated 20,400 search results for *elderly migration*—far more than any other database. (No more than 1,000 records can actually be viewed in GS, however.) AgeLine returned the second most hits (311), but no other database generated more than 300 results. Six of the 12 databases retrieved fewer than 100 records.

Twenty-nine of the 155 relevant articles were each found in just one of the 12 databases. (Eleven unique records were found in AgeLine, nine in GS, and four in SSCI.) Surprisingly, 59 of the relevant articles were not retrieved by any of the 12 databases. This can be attributed to at least four factors: the absence of the terms *elderly* and *migration* within bibliographic records that are nonetheless relevant;[13] the failure to retrieve theoretical and methodological papers that have special relevance for later-life migration but are not themselves *about* later-life migration;[14] the inclusion of key empirical results within papers that deal with multiple age groups or multiple types of migration rather than later-life migration in particular;[15] and the occasional publication of important new information in trade publications that are not indexed by any of the major bibliographic databases.[16]

Recall

As mentioned earlier, recall represents the number of relevant items retrieved as a proportion of all the relevant items that might potentially be retrieved. In this case, it is simply the percentage of the 155 relevant articles retrieved by each of the 12 databases.

When all the search results (up to 300) are considered, Google Scholar and AgeLine outperform the other databases by a wide margin (see table 1).[17] However, no one database retrieves the highest proportion of relevant records within every set of search results ("first 10 hits," "first 20 hits," and so on). The relative effectiveness of each database, therefore, depends on the number of records the searcher is willing to examine. For instance, a searcher willing to examine only the first 10 results will find that MEDLINE, GEOBASE, and Academic Search Elite return the greatest number of relevant articles. If the searcher is willing to evaluate the first 40 hits, then GS is tied for second place, after MEDLINE.

The relative effectiveness of each database, therefore, depends on the number of records the searcher is willing to examine.

Table 1

Recall Rates of Google Scholar and 11 Other Databases (Number of Relevant Records Retrieved as a Percentage of the 155 Relevant Records)

Database	First 10 hits	First 20 hits	First 30 hits[a]	First 40 hits	First 50 hits	First 75 hits[b]	First 100 hits[c]	All search results[d]	N[e]
Google Scholar	4	7	10	12	12	20	25	41	20,400
Academic Search Elite	5	9	11	12	13	15	—	15	73
AgeLine	1	2	3	3	3	5	5	35	311
ArticleFirst	3	7	8	—	—	—	—	8	27
EconLit	3	5	6	6	7	8	8	10	274
GEOBASE	5	7	10	10	12	14	15	15	96
MEDLINE	5	10	13	14	15	16	17	19	174
PAIS International	1	—	—	—	—	—	—	2	12
POPLINE	0	0	1	1	1	4	7	21	295
Social Sciences Abstracts	1	3	6	7	10	—	—	15	57
SSCI	3	7	10	12	13	17	21	26	117
SocINDEX	3	7	10	11	11	13	—	13	91
Rank of Google Scholar	4th	3rd (tied)	3rd (tied)	2nd (tied)	4th (tied)	1st	1st	1st	—

a Data for ArticleFirst are based on all records retrieved—27 records rather than 30.

b Data for Academic Search Elite are based on all records retrieved—73 records rather than 75.

c Data for GEOBASE are based on all records retrieved—96 records rather than 100.

d For Google Scholar, includes the first 300 records retrieved.

e Total number of records retrieved by each search.

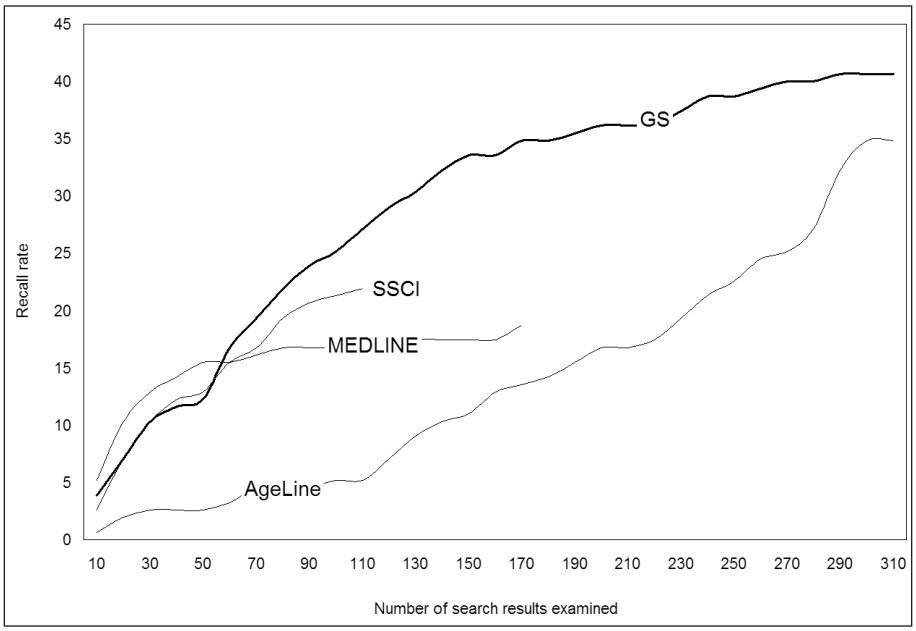

Figure 1. Recall Varies with the Number of Search Results Examined

As table 1 shows, the recall rate of GS puts it within the top four databases regardless of how many search results are examined. At the same time, GS emerges as the single best database only when the searcher is willing to examine more than 56 search results. This can be seen in figure 1, which shows the recall rates of GS, SSCI, MEDLINE, and AgeLine. MEDLINE is best at bringing up relevant articles right at the beginning of the results list. Google Scholar does well within the first 50 hits, although its superior performance is more apparent later in the list of results. With MEDLINE, for example, an examination of hits 50 to 150 will not bring a substantial increase in the number of relevant items found. In contrast, GS continues delivering relevant results up to the 200th hit and beyond. (As shown in figure 1, AgeLine is something of an anomaly, since AgeLine results are sorted by date rather than by relevance.)

These findings reveal that the idiosyncrasies of Google Scholar's search mechanism—the absence of controlled subject terms, for example—do not compromise its ability to retrieve relevant results in response to simple keyword searches. In fact, the GS search mechanism performs better than most. Table 2 shows the number of relevant records retrieved, not as a percentage of all 155 relevant articles but as a percentage of all the relevant articles known to be included in each database. This measure removes the impact of differences in database coverage, thereby highlighting the effectiveness of each database's search mechanism. Table 2 reveals, for instance, that a Google Scholar keyword search for *elderly migration* returns 63 of the 144 relevant articles available within GS (44 percent). Only four of the 12 databases are more effective in retrieving relevant records that are included in the database (and, therefore, potentially retrievable). Together, tables 1 and 2 reveal that the high recall rate of GS can be attributed not

Table 2

Number of Relevant Records Retrieved as a Percentage of the Relevant Records Included in Each Database

Database	Number retrieved	Number included[a]	Percentage retrieved[b]
AgeLine	54	73	74
MEDLINE	29	39	74
POPLINE	32	67	48
EconLit	16	35	46
Google Scholar	63	144	44
GEOBASE	24	60	40
Academic Search Elite	23	62	37
SSCI	40	113	35
SocINDEX	20	64	31
Social Sciences Abstracts	23	86	27
PAIS International	3	18	17
ArticleFirst	12	94	13

[a] Number of relevant records retrieved by a search for *elderly migration*.

[b] Number of the 155 relevant records included in the database. From W.H. Walters, "Google Scholar Coverage of a Multidisciplinary Field," *Information Processing & Management* 43, 4 (2007): 1121–32; W.H. Walters and E.I. Wilder, "Bibliographic Index Coverage of a Multidisciplinary Field," *Journal of the American Society for Information Science and Technology* 54, 14 (2003): 1305–12; and subsequent analyses.

just to its excellent coverage of the literature[18] but also to the effectiveness of its search mechanism.

Precision

Whereas recall represents the effectiveness of each database in retrieving relevant documents, precision indicates how well each database retrieves relevant documents while excluding non-relevant results. Databases with high recall are those that retrieve many relevant records. In contrast, databases with high precision are those for which relevant records make up a high proportion of all the records retrieved. (Specifically, precision is the number of relevant items retrieved as a proportion of all items retrieved.) Early reviewers criticized GS for its apparently low precision but did not support their claims with systematic evidence.[19]

Table 3

Precision of Google Scholar and 11 Other Databases (Number of Relevant Records Retrieved as a Percentage of All Records Retrieved)

Database	First 10 hits	First 20 hits	First 30 hits[a]	First 40 hits	First 50 hits	First 75 hits[b]	First 100 hits[c]	All search results[d]	N[e]
Google Scholar	60	55	53	45	38	41	39	21	20,400
Academic Search Elite	80	70	57	45	40	32	—	32	73
AgeLine	10	15	13	10	8	9	8	17	311
ArticleFirst	50	55	44	—	—	—	—	44	27
EconLit	40	35	30	23	22	16	13	6	274
GEOBASE	70	55	50	40	36	29	25	25	96
MEDLINE	80	80	67	55	48	33	26	17	174
PAIS International	20	—	—	—	—	—	—	25	12
POPLINE	0	0	3	3	4	8	11	11	295
Social Sciences Abstracts	10	20	30	28	32	—	—	40	57
SSCI	40	55	53	48	40	36	33	34	117
SocINDEX	40	55	53	43	34	27	—	22	91
Rank of Google Scholar	4th	3rd (tied)	3rd (tied)	3rd (tied)	4th	1st	1st	8th	—

a Data for ArticleFirst are based on all records retrieved—27 records rather than 30.

b Data for Academic Search Elite are based on all records retrieved—73 records rather than 75.

c Data for GEOBASE are based on all records retrieved—96 records rather than 100.

d For Google Scholar, includes the first 300 records retrieved.

e Total number of records retrieved by each search.

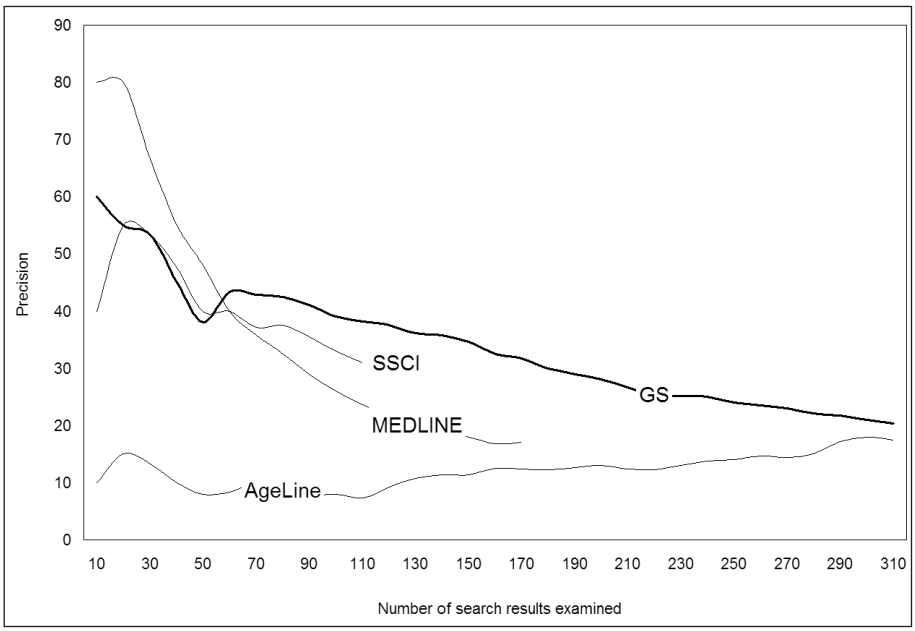

Figure 2. Precision Varies with the Number of Search Results Examined

When the entire set of search results is considered, GS ranks eighth among the 12 databases in terms of precision (see table 3). This is perhaps an unfair comparison, however, since Google Scholar's overall precision (21 percent) reflects its performance over the first 300 search results—far more results than any other database. PAIS International achieves a similar precision score (25 percent) over a set of just 12 search results.

Within the first 20 hits, GS has the third-highest precision of the 12 databases. Specifically, 55 percent of the first 20 records retrieved by GS are relevant. This level of precision is lower than that of MEDLINE (80 percent) and Academic Search Elite (70 percent) but far higher than that of EconLit (35 percent) and Social Sciences Abstracts (20 percent). GS also performs well when 30 or 40 hits are considered, tying for third place in each case.

As shown in figure 2, the precision of Google Scholar remains relatively high even after the first 50 hits. This is its greatest advantage in terms of precision. Within the first 100 search results, 39 percent of GS records but only 26 percent of MEDLINE records are relevant. (As noted earlier, AgeLine records are not sorted by relevance, so the precision of AgeLine does not drop off as additional search results are examined.)

The utility of GS could be improved if relevant results were concentrated more heavily within the first 20 or 30 hits rather than the first 50 or 100.

Figure 2 also reveals that the utility of GS could be improved if relevant results were concentrated more heavily within the first 20 or 30 hits rather than the first 50 or 100. Although highly cited articles are especially likely to appear early in the list of GS search results,[20] there is still room for improvement when Google Scholar's ranking mechanism

is compared with that of MEDLINE. Improvements in the ranking mechanism—or, more specifically, improvements in the mechanism that sorts the top 100 results—may be especially important if the individuals most likely to choose GS are also those least likely to look past the first 10 or 20 hits.

Results of a Title-Only Search

As noted earlier, keyword searches in Google Scholar look for terms not only in bibliographic records and abstracts but also in any indexed full-text content. (Of the 144 relevant records included in the GS database, 25 percent have links to searchable full text.) Despite their superficial similarity to the searches conducted in the other databases, GS keyword searches are more comprehensive.

GS provides no mechanism for limiting the search fields to bibliographic records and abstracts. Moreover, full-text searching is not always available within the other databases. An across-the-board comparison that uniformly excludes or includes bibliographic records, abstracts, and full-text content is, therefore, not possible. All 12 databases do support title-only searching, although that comparison would not accurately represent the behavior of a typical user. (Title searching is not the default option in any of the 12 databases.) However, a comparison of standard searching and title-only searching within GS may help reveal the impact of Google Scholar's full-text search capabilities on its recall and precision. The title search conducted for this purpose was identical to the standard search (see the appendix) except that "in the title of the article" was selected from the drop-down menu labeled "where my words occur."

Because a GS title search retrieves only those documents that have both *elderly* and *migration* in the title, we might expect that limiting the search to the title field would result in lower recall. As table 4 shows, this is indeed the case. The difference in recall is especially significant when more than 50 search results are examined. Even more dramatic is the reduction in the total number of hits, from 20,400 to 127. However, there is virtually no difference in recall within the first 40 or 50 search results. Both the standard search and the title search result in a 12 percent recall rate for the first 40 hits, placing GS in second place among the 10 databases that provide 40 or more search results. Limiting the search to the title field does reduce overall recall, mainly by truncating the results list but also by hindering recall after the first 50 hits.

In terms of precision, the results are much the same (see table 4). Both standard and title searches result in 12 to 13 percent precision over the first 50 hits. Standard searches bring a higher concentration of relevant results when more than 50 hits are examined, although the overall difference in precision is not as great as the overall difference in recall. (Specifically, standard and title searches result in 21 percent and 17 percent precision, respectively, over the set of all search results.)

Although Google Scholar's full-text search capabilities do improve its performance, the gains in both recall and precision occur after the fiftieth hit. For users interested only in the first few dozen search results, a GS search limited to the title field returns the same number and concentration of relevant results as a standard GS search. Moreover, these results suggest that Google Scholar would perform relatively well even if it did not search the full text of each document for which full text is available.

Table 4

Recall and Precision of Google Scholar Standard and Title-Only Searches

	First 10 hits	First 20 hits	First 30 hits	First 40 hits	First 50 hits	First 75 hits	First 100 hits	All search results[a]	N[b]
Recall									
Standard search	4	7	10	12	12	20	25	41	20,400
Title-only search	3	7	10	12	13	14	14	14	127
Rank of standard search	4th	3rd	3rd	2nd	4th	1st	1st	1st	—
Rank of title-only search	4th	3rd	3rd	2nd	2nd	4th	4th	8th	—
Precision									
Standard search	60	55	53	45	38	41	39	21	20,400
Title-only search	50	55	53	45	40	28	22	17	127
Rank of standard search	4th	3rd	3rd	3rd	4th	1st	1st	8th	—
Rank of title-only search	4th	3rd	3rd	3rd	2nd	5th	4th	8th	—

a Includes the first 300 records retrieved.

b Total number of records retrieved by each search.

Conclusions

This study addresses Google Scholar's search performance within one particular subject area: later-life migration. Because database performance varies considerably from one field to another,[21] evaluations based on other search topics might yield different results. Nonetheless, these findings suggest that for at least some topics, GS performs better than many subscription databases.

The high recall rate of Google Scholar is consistent with its excellent coverage of the later-life migration literature. GS includes records for more than 90 percent of the relevant documents[22] and consequently retrieves a greater number of relevant results than the other 11 databases. Perhaps more surprising is the high precision of GS. Although early reviews of Google Scholar noted its apparently low precision, GS consistently ranks among the top four databases when the first 10 to 100 search results are examined.

These findings suggest that a searcher who is unwilling to search multiple databases or to adopt a sophisticated search strategy is likely to achieve better than average recall and precision by using Google Scholar.

These findings suggest that a searcher who is unwilling to search multiple databases or to adopt a sophisticated search strategy is likely to achieve better than average recall and precision by using Google Scholar. Of course, there may be other reasons for preferring conventional databases, such as the need to develop and practice advanced searching skills, either for use in later research or as a means of encouraging critical thinking and conceptual clarity in academic work.[23]

Evaluating Relevance in the Educational Setting

Several features of GS are likely to make it especially attractive to college and university students. In particular, the GS search interface conforms to the expectations that many searchers have developed through their use of Google and other Web search engines.[24] Research by Bernard J. Jansen, Amanda Spink, and others shows that most Web searchers conduct simple searches, then examine relatively few records. Approximately 25 percent of all Web search queries consist of just a single term, and fewer than 20 percent include a Boolean operator.[25]

The standard of relevance used in this study may be especially appropriate in college or university settings. The approach adopted here, based on the expert evaluation of complete articles rather than citations or abstracts, contrasts with those methods that rely on bibliometric relationships or on users' own assessments of relevance.[26] The databases that perform best in this analysis are those that consistently lead users to documents that have met comparatively strict standards for relevance of topic, importance of findings, and innovativeness of methods or approach.[27] Arguably, these are the documents that students ought to read in order to achieve a good understanding of the subject.

Although relevance can be defined in many ways, it is nearly always understood in terms of the information seeker's needs or desires.[28] In the academic setting, where the

information seeker is most often a student, we can identify a second kind of relevance—relevance to the educational goals of the instructor. By that standard, relevance refers not just to the document characteristics most important to the student (topic, novelty, readability, authority, length, and so on) but to a set of more general educational expectations. If, for example, instructors believe that students benefit from reading high-quality writing and analysis, they may favor online databases and search mechanisms that maximize students' likelihood of retrieving high-quality documents—documents selected not just for their relevance to a particular task or assignment but also for their value as examples of good scholarly work.

Under this standard of relevance, quality may be defined using whatever criteria suit the instructor's purposes. Scholarly impact, pedagogical value, clarity of presentation, historical importance, strength of argument, emotional impact, and breadth of practical application might each be given priority in different contexts. Because students do not always have the expertise needed to judge the quality of their search results in these terms,[29] instructors and librarians may want to adopt strategies that increase students' exposure to high-quality research by (1) encouraging the use of print and online collections that have adopted rigorous collection development standards (JSTOR, for example) and (2) favoring databases such as GS that maximize the number of high-quality documents and minimize the number of low-quality documents retrieved.

This general perspective on relevance can also be applied to other audiences (high school students, hospital patients) and other kinds of information resources (statistical databases, collections of literary works, business resources, and so on). Relevance judgments that account for third-party assessments of quality may be especially appropriate whenever the ultimate goals of the institution extend beyond the provision of task-specific information.

Acknowledgements

I am grateful for the comments of Esther Isabelle Wilder and two anonymous referees.

William H. Walters is dean of library services and associate professor of social sciences, Menlo College, Atherton, CA; he may be contacted via e-mail at: wwalters@menlo.edu.

Appendix

Database Search Procedures

All searches were conducted in February 2008. Each database covers the entire period in which relevant documents were published (January 1990 through December 2000).

Google Scholar

Platform: Google Scholar Web interface, http://scholar.google.com/.
User behavior: Typing *elderly migration* in the search box of the basic search interface.
Actual search (to account for the publication dates of the relevant documents): Typed *elderly migration* in the "with all of the words" search box of the advanced scholar search interface. Used the date selection boxes.
Fields searched: All fields of the bibliographic record, abstract, and full text. All the full-text content available to GS is indexed even when that content cannot be viewed by the user due to licensing restrictions. Consequently, the search results do not vary in response to differences in institutional library holdings.
Records retrieved: All records that have both *elderly* and *migration*. The two words need not appear near each other or in that order.
Results were sorted by relevance.

Academic Search Elite

Platform: EBSCOhost.
User behavior: Typing *elderly and migration* in the search box of the basic search interface.
Actual search (to account for the publication dates of the relevant documents): Typed *elderly and migration* in the search box of the basic search interface. Used the published date selection boxes.
Fields searched: Author, subject, keyword, article title, source title, abstract.
Records retrieved: All records that have both *elderly* and *migration*. The two words need not appear near each other or in that order.
Results were sorted by relevance.

AgeLine

Platform: AgeLine Web interface, http://www.aarp.org/research/ageline/.
User behavior: Typing *elderly migration* in the search box of the basic search interface.
Actual search (to account for the publication dates of the relevant documents): Typed *elderly migration* in the search box of the basic search interface. Used the year selection boxes.
Fields searched: All fields of the bibliographic record and abstract.

Records retrieved: All records that have the character string *migrati*, since *elderly* is a stop word in AgeLine, and all search terms of more than seven letters are automatically truncated after the seventh letter.

Results were sorted by date (most recent first). Relevance sorting is not available in AgeLine.

ArticleFirst

Platform: OCLC FirstSearch.

User behavior: Typing *elderly migration* in the search box of the advanced search interface and selecting keyword as the search field. Selecting relevance ranking.

Actual search (to account for the publication dates of the relevant documents): Typed *elderly migration* in the search box of the advanced search interface and selected keyword as the search field. Selected relevance ranking and used the year selection box.

Fields searched: Title, subject heading, notes. (ArticleFirst records do not have abstracts.)

Records retrieved: All records that have both *elderly* and *migration*. The two words need not appear near each other or in that order.

Results were sorted by relevance.

EconLit

Platform: EBSCOhost.

User behavior: Typing *elderly and migration* in the search box of the basic search interface.

Actual search (to account for the publication dates of the relevant documents): Typed *elderly and migration* in the search box of the basic search interface. Used the published date selection boxes.

Fields searched: Author, subject, keyword, article title, source title, abstract.

Records retrieved: All records that have both *elderly* and *migration*. The two words need not appear near each other or in that order.

Results were sorted by relevance.

GEOBASE

Platform: OCLC FirstSearch.

User behavior: Typing *elderly migration* in the search box of the advanced search interface and selecting keyword as the search field. Selecting relevance ranking.

Actual search (to account for the publication dates of the relevant documents): Typed *elderly migration* in the search box of the advanced search interface and selected keyword as the search field. Selected relevance ranking and used the year selection box.

Fields searched: Title, subject heading, abstract, and notes.

Records retrieved: All records that have both *elderly* and *migration*. The two words need not appear near each other or in that order.

Results were sorted by relevance.

MEDLINE

Platform: OCLC FirstSearch.
User behavior: Typing *elderly migration* in the search box of the advanced search interface and selecting keyword as the search field. Selecting relevance ranking.
Actual search (to account for the publication dates of the relevant documents): Typed *elderly migration* in the search box of the advanced search interface and selected keyword as the search field. Selected relevance ranking and used the year selection box.
Fields searched: Title, subject heading, abstract, and notes.
Records retrieved: All records that have both *elderly* and *migration*. The two words need not appear near each other or in that order.
Results were sorted by relevance.

PAIS International

Platform: CSA Illumina.
User behavior: Typing *elderly and migration* in the search box of the quick search interface.
Actual search (to account for the publication dates of the relevant documents): Typed *elderly and migration* in a single search box of the advanced search interface. Selected anywhere as the search field. Used the date range selection boxes.
Fields searched: All fields of the bibliographic record and abstract.
Records retrieved: All records that have both *elderly* and *migration*. The two words need not appear near each other or in that order.
Results were sorted by relevance.

POPLINE

Platform: POPLINE Web interface, http://db.jhuccp.org/ics-wpd/popWeb /.
User behavior: Typing *elderly & migration* in the subject search box of the basic search interface.
Actual search (to account for the publication dates of the relevant documents): Typed *elderly & migration* in the subject search box of the advanced search interface. Used the year selection box. Conducted 11 searches, one for each year, since the year selection box does not permit the selection of multiple years.
Fields searched: All fields of the bibliographic record and abstract.
Records retrieved: All records that have both *elderly* and *migration*. The two words need not appear near each other or in that order.
Results were sorted by date (most recent first). Relevance sorting is not available in POPLINE.

Social Sciences Abstracts

Platform: WilsonWeb.
User behavior: Typing *elderly and migration* in a single search box of the advanced search interface. Selecting keyword as the search field.

Actual search (to account for the publication dates of the relevant documents): Typed *elderly and migration* in a single search box of the advanced search interface. Selected keyword as the search field. Used the limit dates selection boxes.
Fields searched: All fields of the bibliographic record and abstract.
Records retrieved: All records that have both *elderly* and *migration*. The two words need not appear near each other or in that order.
Results were sorted by date (most recent first). Relevance sorting is available in Social Sciences Abstracts, but all results have 100 percent relevance when keyword searching is used.

Social Sciences Citation Index

Platform: Web of Science.
User behavior: Typed *elderly migration* in the Web of Science search box. Selected topic as the search field and SSCI as the database.
Actual search (to account for the publication dates of the relevant documents): Typed *elderly migration* in the Web of Science search box. Selected topic as the search field and SSCI as the database. Used the time span selection boxes.
Fields searched: All fields of the bibliographic record and abstract. (A general search—not a cited reference search.)
Records retrieved: All records that have both *elderly* and *migration*. The two words need not appear near each other or in that order.
Results were sorted by relevance.

SocINDEX

Platform: EBSCOhost.
User behavior: Typing *elderly and migration* in the search box of the basic search interface.
Actual search (to account for the publication dates of the relevant documents): Typed *elderly and migration* in the search box of the basic search interface. Used the published date selection boxes.
Fields searched: Author, subject, keyword, article title, source title, abstract.
Records retrieved: All records that have both *elderly* and *migration*. The two words need not appear near each other or in that order.
Results were sorted by relevance.

Notes

1. G.E. Gorman, "Giving Way to Google," *Online Information Review* 30, 2 (2006): 97–9; Benjamin P. Norris, "Google: Its Impact on the Library," *Library Hi Tech News* 23, 9 (2006): 9–11; Andrew K. Pace, "If You Can't Beat 'em, Join 'em," *American Libraries* 36, 8 (2005): 78–9; Jeffrey Pomerantz, "Google Scholar and 100 Percent Availability of Information," *Information Technology and Libraries* 25, 2 (2006): 52–6; and Carol Tenopir, "Remaining Relevant Online," *Library Journal* 132, 10 (2007): 32.
2. Laura Bowering Mullen and Karen A. Hartman, "Google Scholar and the Library Web Site: The Early Response by ARL Libraries," *College & Research Libraries* 67, 2 (2006): 106–22;

Chris Neuhaus, Ellen Neuhaus, and Alan Asher, "Google Scholar Goes to School: The Presence of Google Scholar on College and University Web Sites," *Journal of Academic Librarianship* 34, 1 (2008): 39–51.

3. Janet Adlington and Chris Benda, "Checking Under the Hood: Evaluating Google Scholar for Reference Use," *Internet Reference Services Quarterly* 10, 3/4 (2005): 135–48; Rebecca Donlan and Rachel Cooke, "Running With the Devil: Accessing Library-Licensed Full-Text Holdings Through Google Scholar," *Internet Reference Services Quarterly* 10, 3/4 (2005): 149–57; Peter Jacsó, "As We May Search: Comparison of Major Features of the Web of Science, Scopus, and Google Scholar Citation-Based and Citation-Enhanced Databases," *Current Science* 89, 9 (2005): 1537–47, http://www.iisc.ernet.in/currsci/nov102005/1537.pdf (accessed September 28, 2008); Jacsó, "Google Scholar: The Pros and the Cons," *Online Information Review* 29, 2 (2005): 208–14; Jacsó, "Deflated, Inflated and Phantom Citation Counts," *Online Information Review* 30, 3 (2006): 297–309; Philipp Mayr and Anne-Kathrin Walter, "An Exploratory Study of Google Scholar," *Online Information Review* 31, 6 (2007): 814–30; Martin Myhill, "Google Scholar," *Charleston Advisor* 6, 4 (2005): 49–52; Mick O'Leary, "Google Scholar: What's In It For You?" *Information Today* 22, 7 (2005): 35–9; and Joann M. Wleklinski, "Studying Google Scholar: Wall to Wall Coverage?" *Online* 29, 3 (2005): 22–6.
4. Burton Callicott and Debbie Vaughn, "Google Scholar vs. Library Scholar: Testing the Performance of Schoogle," *Internet Reference Services Quarterly* 10, 3/4 (2005): 82.
5. D.Yvonne Jones, "Biology Article Retrieval From Various Databases: Making Good Choices with Limited Resources," *Issues in Science & Technology Librarianship* 44 (Fall 2005), http://www.istl.org/05-fall/refereed.html (accessed September 27, 2008).
6. Susan Gardner and Susanna Eng, "Gaga over Google? Scholar in the Social Sciences," *Library Hi Tech News* 22, 8 (2005): 43.
7. Rena Helms-Park, Pavlina Radia, and Paul Stapleton, "A Preliminary Assessment of Google Scholar as a Source of EAP Students' Research Materials," *Internet and Higher Education* 10, 1 (2007): 65–76.
8. William H. Walters, "Google Scholar Coverage of a Multidisciplinary Field," *Information Processing & Management* 43, 4 (2007): 1121–32.
9. For details, see Walters, "Later-Life Migration in the United States: A Review of Recent Research," *Journal of Planning Literature* 17, 1 (2002): 37–66. The 155 relevant documents include all the works cited in that review, except for those published before 1990 or after 2000; those published as books, book chapters, or dissertations; and those that are primarily bibliographic or editorial in nature (10 items).
10. All the full-text content available to GS is indexed even when that content cannot be viewed by the user due to licensing restrictions. Consequently, the GS search results do not vary in response to differences in institutional library holdings. According to a 2004 investigation, early versions of GS indexed only the first few pages of long full-text documents. See Gary Price, "Google Scholar Documentation and Large PDF Files," SearchEngineWatch.com (December 1, 2004), http://blog.searchenginewatch.com/blog/041201-105511 (accessed September 28, 2008). It is not clear whether GS now indexes long documents in their entirety.
11. Marian Burright, "Google Scholar: Science & Technology," *Issues in Science & Technology Librarianship* 45 (Winter 2006), http://www.istl.org/06-winter/databases2.html (accessed September 28, 2008); Gardner and Eng; Chris Neuhaus et al., "The Depth and Breadth of Google Scholar: An Empirical Study," *portal: Libraries and the Academy* 6, 2 (2006): 127–41; Roy Tennant, "Google, the Naked Emperor," *Library Journal* 130, 13 (2005): 29; Tennant, "Is Metasearching Dead?" *Library Journal* 130, 12 (2005): 28; and Mayr and Walter.
12. Judit Bar-Ilan, "An Ego-Centric Citation Analysis of the Works of Michael O. Rabin Based on Multiple Citation Indexes," *Information Processing & Management* 42, 6 (2006): 1553–66; Kathleen Bauer and Nisa Bakkalbasi, "An Examination of Citation Counts in a New Scholarly Communication Environment," *D-Lib Magazine* 11, 9 (2005), http://www.dlib.

org/dlib/september05/bauer/09bauer.html (accessed September 28, 2008); Jacsó, "As We May Search"; Jacsó, "Comparison and Analysis of the Citedness Scores in Web of Science and Google Scholar," in *Digital Libraries: Implementing Strategies and Sharing Experiences*, ed. Edward A. Fox et al. (New York: Springer, 2005), 360–9; and Jacsó, "Deflated, Inflated and Phantom Citation Counts."

13. Albert Chevan, "Holding On and Letting Go: Residential Mobility During Widowhood," *Research on Aging* 17, 3 (1995): 278–302; Kevin E. McHugh and Robert C. Mings, "On the Road Again: Seasonal Migration to a Sunbelt Metropolis," *Urban Geography* 12, 1 (1991): 1–18.
14. Gary L. Hunt, "Equilibrium and Disequilibrium in Migration Modeling," *Regional Studies* 27, 4 (1993): 341–9; Andrei Rogers and Alain Belanger, "The Importance of Place of Birth in Migration and Population Redistribution Analysis," *Environment and Planning A* 22, 2 (1990): 193–210.
15. David E. Clark and William J. Hunter, "The Impact of Economic Opportunity, Amenities and Fiscal Factors on Age-Specific Migration Rates," *Journal of Regional Science* 32, 3 (1992): 349–65; John A. Fulton, Glenn V. Fuguitt, and Richard M. Gibson, "Recent Changes in Metropolitan-Nonmetropolitan Migration Streams," *Rural Sociology* 62, 3 (1997): 363–84.
16. Timothy D. Hogan and Stephen K. Happel, "1993–94 Winter Residents Important to AZ Economy," *Arizona Business* 41, 7 (1994): 1–4.
17. An earlier study (Walters, "Google Scholar Coverage") showed that 45 of the 144 relevant articles indexed by GS were included in Google Scholar only due to their appearance in the bibliographies of papers previously indexed by GS. Because all 155 relevant articles can be found in the bibliography of one particular paper (Walters, "Later-Life Migration"), it is conceivable that some relevant articles appear in GS solely because of their inclusion in that bibliography. That is, the publication of "Later-Life Migration" might have artificially inflated the number of relevant articles found by the GS keyword search undertaken for this analysis. However, further investigation revealed that of the 45 relevant articles taken by GS from the bibliographies of previously indexed papers, 41 had appeared in the bibliographies of one or more GS-indexed articles published prior to "Later-Life Migration." Of the four relevant articles that might have been indexed solely due to their inclusion in the bibliography of "Later-Life Migration," only two are returned by a GS keyword search for *elderly migration*.
18. Walters, "Google Scholar Coverage."
19. Gardner and Eng; Myhill; O'Leary; and Wleklinski.
20. Marilyn Christianson, "Ecology Articles in Google Scholar: Levels of Access to Articles in Core Journals," *Issues in Science & Technology Librarianship* 49 (Winter 2007), http://www.istl.org/07-winter/refereed.html (accessed September 28, 2008).
21. See, for example, Chris Buckley and Janet Walz, "The TREC-8 Query Track," in *Information Technology: The Eighth Text Retrieval Conference (TREC-8)*, ed. Ellen M. Voorhees and Donna K. Harman (Gaithersburg, MD: U.S. Department of Commerce, National Institute of Standards and Technology, 2000), 65–75, http://trec.nist.gov/pubs/trec8/papers/qtrack.pdf (accessed July 26, 2008).
22. Walters, "Google Scholar Coverage."
23. Because most academic libraries have educational missions that extend beyond the provision of information, the most effective databases for information retrieval are not necessarily the best for educational purposes. See Veronica Calderhead, "Reflections on Information Confusion in Chemistry Information Learning: The Meaning of the Shift from Library Instruction to Information Literacy," *Research Strategies* 16, 4 (1998): 285–99; and Steven J. Herro, "Bibliographic Instruction and Critical Thinking," *Journal of Adolescent & Adult Literacy* 43, 6 (2000): 554–8.
24. Glenn Haya, Else Nygren, and Wilhelm Widmark, "Metalib and Google Scholar: A User Study," *Online Information Review* 31, 3 (2007): 365–75.

25. See 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, 1 (2006): 248–63; Bernard J. Jansen, Amanda Spink, and Tefko Saracevic, "Real Life, Real Users, and Real Needs: A Study and Analysis of User Queries on the Web," *Information Processing & Management* 36, 2 (2000): 207–27; Spink et al., "Searching the Web: The Public and Their Queries," *Journal of the American Society for Information Science and Technology* 52, 3 (2001): 226–34; and Dietmar Wolfram et al., "Vox Populi: The Public Searching of the Web," *Journal of the American Society for Information Science and Technology* 52, 12 (2001): 1073–4. The search engines evaluated by these authors did not include Google or GS.
26. Anselm Spoerri, "Using the Structure of Overlap Between Search Results to Rank Retrieval Systems Without Relevance Judgments," *Information Processing & Management* 43, 4 (2007): 1059–70.
27. Walters, "Later-Life Migration."
28. Pia Borlund, "The Concept of Relevance in IR," *Journal of the American Society for Information Science and Technology* 54, 10 (2003): 913–25; Stefano Mizzaro, "Relevance: The Whole History," *Journal of the American Society for Information Science* 48, 9 (1997): 810–32; Tefko Saracevic, "Relevance: A Review of and a Framework for the Thinking on the Notion in Information Science," *Journal of the American Society for Information Science* 26, 6 (1975): 321–43; Saracevic, "Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science; Part II: Nature and Manifestations of Relevance," *Journal of the American Society for Information Science and Technology* 58, 13 (2007): 1915–33; Saracevic, "Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science; Part III: Behavior and Effects of Relevance," *Journal of the American Society for Information Science and Technology* 58, 13 (2007): 2126–44; and Arthur R. Taylor et al., "Relationships Between Categories of Relevance Criteria and Stage in Task Completion," *Information Processing & Management* 43, 4 (2007): 1071–84.
29. Walters, "Expertise and Evidence in the Assessment of Library Service Quality," *Performance Measurement and Metrics* 4, 3 (2003): 98–102.