ARTICLE

Irrelevant Discovery Layers?

An Evidence-Based Evaluation of Three Common Library Search Tools

Ruth Szpunar, Eric Bradley, Erin Gabrielson, and Catherine Pellegrino

ABSTRACT

Over the past fifteen years, most academic librarians have implemented one-stop search tools, commonly referred to as discovery layers, to accommodate contemporary user expectations. In more recent years these tools have come under criticism due to their limitations and shortcomings. We set out to evaluate if a discovery layer, when prompted with typical user keyword search strings, produced the most relevant search results when compared with two other widely accessible academic search tools. We compared search results from a discovery layer with a central index (WorldCat Discovery) with search results from a subscription interdisciplinary index and abstract database (Academic Search Complete) and a freely accessible academic web search engine (Google Scholar). We created a rubric detailed enough for multiple evaluators, who were the authors, to judge search results for currency, relevancy, proximity, and authority, as well as to assign appropriate penalties. Academic Search Complete search results received the highest overall scores, while WorldCat Discovery search results received the lowest overall scores. When considering individual pieces of the rubric, Academic Search Complete provided the most current and authoritative sources, while Google Scholar provided the most relevant sources. This article provides recommendations for libraries moving forward to consider the benefits and costs of discovery tools.

INTRODUCTION

Over the past 20 years, libraries have been purchasing or developing one-stop search tools that combine the majority of their print and electronic holdings into one, ideally easy to use, search location. This pursuit was encouraged in part by a 2005 OCLC report which found that "Library Web sites were selected by just 1 percent of respondents as the source used to begin an information search."[1] In comparison, 84 percent of these respondents started at a search engine. Libraries, along with their commercial vendors, have sought to provide leading edge one-stop search functionality for their patrons through tools such as federated searches and central index discovery layers in a frantic attempt to keep up with Google as well as its free academic search engine Google Scholar. Two decades into this race, have libraries or library vendors succeeded at this dream, or is it time to consider other strategies to best support library users with their information finding needs?

About the Authors

Ruth Szpunar (ruth@palni.edu) (corresponding author) is Reference, Instruction, and Outreach (RIO) Coordinator and Consultant, PALNI. **Eric Bradley** (ebradley@goshen.edu) is Reference, Instruction, and Outreach (RIO) Coordinator, PALNI and Head Librarian of the Mennonite Historical Library, Goshen College. **Erin Gabrielson** (egabrielson@franklincollege.edu) is Student Learning & Research Librarian, Franklin College. **Catherine Pellegrino** (cpellegr@saintmarys.edu) is Reference Librarian, Saint Mary's College.

Submitted: 17 October 2024. Accepted for Publication: 21 February 2025. Published: 16 June 2025.

https://doi.org/10.5860/ital.v44i2.17266

This study sought to evaluate whether a discovery layer, when prompted with typical user keyword search strings, produced the most relevant search results when compared with two other widely accessible search tools. The authors compared three contemporary means with which users find academic resources at one small liberal arts college: a central index discovery layer, a subscription interdisciplinary index and abstract database, and a freely accessible academic web search engine. Four librarians evaluated the first five results of search strings pulled from the college library's discovery layer analytics, using a shared rubric that ranked currency, relevancy, proximity, and authority, as well as marking penalties. Our findings demonstrated significant shortcomings for the discovery layer as compared with its two peers. In an era of declining enrollments and financial constraints, libraries need to consider seriously whether the investment of the discovery layer is feasible moving forward.

LITERATURE REVIEW

Librarians and scholars began evaluating and comparing search results of federated searches and central index discovery layers (including both open web and commercial library search products) almost immediately upon their arrival in the marketplace. Scholars began to compare Google Scholar to subject-specific databases and library catalogs while Google's tool was still in beta release in 2004.[2] Likewise, researchers compared federated search tools and discovery tools when they were implemented by early adopters as soon as 2006.[3] This literature review focuses on research that evaluates the search results of academic discovery search tools as well as writing that considers the current goals and objectives of these tools.

Research that assesses search tools must define specific criteria for evaluation. While each study in this review selected different criteria and used different language to describe those criteria, the most common elements that were evaluated included availability, coverage, currency, quality, relevance, and volume. Quantitative approaches were used for each of these criteria except for relevance and quality. Most articles evaluated relevance, although several limited their evaluation to exclusively quantitative criteria.[4] Studies that examined relevance and quality relied on subjective evaluations by appointed expert(s), evaluated search results against established lists, or followed set criteria for evaluation.

The Challenge of Evaluating Relevance

For multiple reasons, it is common in the literature to evaluate the relevance of topic searches subjectively. Pulikowski and Matysek argue that because users evaluate search results subjectively, it is appropriate for a research study to do so also.[5] Ciccone and Vickery, as well as Noe, point to the challenge of not knowing the intent of the searcher and conclude that subjective evaluation is necessary.[6] Dahlen and Hanson also highlight this limitation, but created a scoring rubric.[7] Many of the studies using subjective evaluation seek out subject experts to conduct evaluation for relevancy or quality. Lee and Chung, in addition to Zhang, utilized graduate students to evaluate searches, while other studies utilized reference or subject librarians who were not the study investigators.[8] In other studies, the subjective evaluation was conducted by the authors.[9] All of the studies that employed subjective evaluation examined topic searches, except for Ciccone and Vickery, who also included known item searches.[10]

One approach to mitigate problems with subjective evaluation of search results is to develop or use an established list of articles and then evaluate search results against this list. In multiple studies Walters compared search results to a personally compiled list.[11] This list of 155 relevant articles in one subject area was "identified through database searching, citation tracing, journal

browsing, and consultation with colleagues in the social sciences."[12] Jones compared search results to coverage in one subject-specific database.[13] Additional studies limit their analysis to known item searches to determine relevance, using usage reports or bibliographic citations.[14]

Another approach to provide a more objective evaluation of search result relevance is the use of scoring rubrics for topic searches. Scoring rubrics for topic searches take several different forms in the literature. Brophy and Bawden, along with Howland et al. evaluated specific criteria in rubric form, but did not provide guidelines on how to assess the score for each criterion.[15] Scoring rubrics were developed by Asher et al., Dahlen and Hanson, and Georgas, although only Dahlen and Hanson provided guidelines for determining relevance and did so in broad strokes.[16] The clearest scoring rubric guidelines for evaluating relevance were developed by Gardner and Eng, as well as Cooke and Donlan.[17] Gardner and Eng assigned various points based on whether the exact search terms appeared within the title, abstract, full text or descriptors.[18] Cooke and Donlan assigned points on whether exact search terms or synonyms appeared in the title.[19]

The large majority of studies that included Google or Google Scholar and evaluated relevance as a stand-alone criterion (9 out of 13) found Google's tools to provide the most relevant results compared to other search tools.[20] When Google was compared against Google Scholar for relevance, Google results were found most relevant each time.[21] However, when quality of results was included as a factor in the relevance scores, EBSCO Discovery Service was found to have the best results.[22] Also, in studies that evaluated currency as a stand-alone criterion, Google Scholar never had the newest material; instead, subject-specific databases featured the most current material.[23] In studies evaluating coverage as a stand-alone criterion, Google Scholar consistently had the greatest content coverage.[24] Studies that did not include Google or Google Scholar and evaluated relevance as a stand-alone criterion had varied results, but much of this was due to the wide variety of tools evaluated.

Although outside the scope of this literature review, usability studies and bibliometric analyses offer additional evaluative research on federated searches and central index discovery layers. Wisneski compiled a robust review of the literature on usability testing of discovery layers, while Hoeve and Geuther, as well as Chen, provide examples of using bibliometric analysis with discovery tools.[25]

Google Scholar and Library Search Tools

In recent years, a growing number of researchers are questioning, as we are, the goals of central index discovery layers and their success at meeting those goals. Google Scholar and other open academic search tools are receiving increasing criticism for their limitations from outside the library literature. Achenbach et al. criticize Google Scholar for focusing primarily on publications, well-cited papers, English content, and discipline-specific resources.[26] Relevancy biases have been noted with Google Scholar regarding English content and highly cited content.[27] Von Hippel and Buck call to task Google Scholar's emphasis on selective journals, citations, and author prestige which they argue overrepresents men at prestigious universities in the United States, calling that overrepresentation "poor proxies for research quality."[28] Redkina, as well as Khalid et al., also points out that Google Scholar does not exhaustively search the academic literature.[29] Khalid et al. also criticize Google Scholar's reliance on citation count as a factor in determining relevance and point out that Google Scholar sometimes deems poor research as relevant because it is cited by multiple authors.[30]

A variety of researchers within the library literature have criticized the goals and workings of discovery layers. Wisneski argues that the goal of discovery layers may be impossible to achieve for libraries, instead arguing that librarians should work with students in using discovery tools to help students learn their construction along with their strengths and weaknesses.[31] Many of the studies in this literature review raise concerns about the lack of transparency from Google and library commercial vendors with proprietary search algorithms.[32] More recently, Eller calls out Google's "black box algorithm" and calls for library discovery systems to provide transparency in the mechanics of their search algorithms so that library staff are well equipped when creating metadata.[33] The growing incorporation of artificial intelligence into search tools through machine learning algorithms adds complexity to the issue; even when the workings of algorithms in academic search tools are fully transparent there can be an inability to explain search results.[34]

The current study fills a gap in the existing literature by evaluating the relevancy of search results using a scoring rubric that is detailed enough to be used by multiple evaluators and includes multiple facets of relevance. While many previous studies identified the challenges of determining relevancy, especially when interpreting another person's search queries, each discovery tool is using an internal, and often hidden, mechanism for determining relevance for end users. We believe that starting a public conversation for libraries and vendors about what determines a search result's relevance is important.

SEARCH TOOL OVERVIEW

For this study, we evaluated search results in both print and electronic formats from three products: a discovery layer with a central index (WorldCat Discovery), a subscription interdisciplinary index and abstract database (Academic Search Complete), and a freely accessible academic web search engine (Google Scholar).

WorldCat Discovery is the current interface for OCLC's WorldCat product, launched in 2014.[35] This tool utilizes a central index including OCLC's WorldCat and FirstSearch proprietary databases along with resources from other database providers and publishers with title lists provided.[36] OCLC has heavily developed the product since its launch. Recently, the developers updated their search backend to use Apache Solr, which allowed them to further refine the algorithm.[37] Individual users can sort their search results using a variety of options; within the Private Academic Library Network of Indiana (PALNI) consortium most libraries default to either Best Match or to Library. Best Match prioritizes search terms found in title, author, subject, and abstract fields as well as newer and locally held material, while Library prioritizes materials that are available to the user in local and consortial collections as shown in Table 1.[38]

Table 1. WorldCat Discovery Sort Options

Best Match	"Blends boosts on Title, Author, Subject and Abstract, while considering word order and position. Prefers recent over older materials and popular over rare materials with boosts on Publication Date and Worldwide Holdings. Prefers materials from local and group collections, boosting those over materials from other libraries worldwide."[39]
Library	"Shares all features of the Best Match sort except that it dramatically boosts materials from local and group collections. Library demands local collections over group and group over worldwide."[40]

Academic Search Complete is one iteration of EBSCO's Academic Search proprietary line of journal and magazine aggregator databases originally launched in 1997.[41] EBSCO shares what materials are indexed in Academic Search Complete in title lists and also discloses some details on how their search algorithm works. EBSCO reports that the main element in determining relevance is how often a user's search term matches the metadata and full-text from a record, prioritizing newer and peer-reviewed material while deemphasizing book reviews and news material. They also include consideration for phrase adjacency in their algorithm.[42]

Google Scholar, Google's search product for academic literature, has been freely available since 2004. Google claims to "rank documents the way researchers do" based on publication location, author, and citation counts.[43] While library and information science literature about Google Scholar often refers to both its search code and ranking algorithms as an indecipherable "black box," researchers have used reverse engineering to determine some facts about what Google Scholar searches and how it ranks results. Marks and Le determined that Google Scholar searches and indexes articles from familiar databases, including but not limited to: HeinOnline, JSTOR, SciELO, SSRN, ProQuest, Wiley, EBSCO, Elsevier, Sage, and LexisNexis.[44] Google Scholar's own Help page explains how individuals can make sure their work is indexed by merely uploading their paper as a PDF hosted on a .edu website, including a bibliography or works cited, and including the title and author names on the first page of the PDF. Google Scholar also implies that its search crawls university repositories, as they offer specific tips for repository access, and suggest specific repository software (Eprints, Digital Commons, or DSpace). The Google Scholar Help page also makes it clear what smaller journal publishers can do to ensure their publications appear in searches, thus confirming that Google Scholar does crawl "established journal hosting services" and journal aggregators like JSTOR.[45] Overall, this amounts to a vast collection of books, articles, and other scholarly material that Google Scholar searches.

Google Scholar is not in the business of creating metadata about the results it finds. Metadata for what Google Scholar searches would of course be inconsistent from one website it crawls to another, leaving the user to discover on their own whether a particular result is from a journal that uses peer-review, for instance, or if the result is a dissertation. The lack of filters and limiters in Google Scholar is understandable considering how it searches across varied aggregators, repositories, and .edu websites.

Other researchers have used reverse engineering to gain insights into the factors that Google Scholar weighs in its results display. Beel and Gipp explored how much citation counts matter in the Google Scholar ranking algorithm.[46] They had previously established that Google Scholar gives higher ranking to results that have exact search query terms in the title and that are newer. They ultimately confirmed that higher citation counts rank the result higher, but they were unable to conclude why Google Scholar sometimes makes exceptions to the citation count weight when searching in full text. Ultimately, Beel and Gipp provide a succinct explanation for one possible reason why Google Scholar is so popular among undergraduate researchers: "... due to the strong weight on citation counts, Google Scholar is more suited when searching for popular standard literature than for searching gems, the latest trends, or papers whose authors are advancing views opposite to the mainstream."[47]

METHODS

For this study, we used four categories for ranking each search result. We started with three categories: currency, relevancy, and authority, based on our literature review, focusing on the

work of Howland et al. and Georgas.[48] As we evaluated our first set of results, we found that specific keyword phrases were not receiving enough value, so we added a proximity category to counter this. We also created a penalty category to address problematic elements of some results. The full rubric is included in appendix A.

For currency, items that were more than 50 years old received a score of 0, items that were more than 10 years old received a 1, items published within the past 6–10 years received a 2, and items published within the last 5 years received a 3. We assigned these scores automatically, using Google Sheets formulas to assess each search result. We did not adjust the currency rubric according to the discipline of the search query, because of the subjectivity of assigning a discipline to many of the search terms. Given the amount of subjectivity that already exists in the literature on evaluating search results, we wanted to avoid that in our study, while acknowledging the limitations that arise from this approach.

Relevancy was determined based on the number of individual words from the original search that appeared in the search result, with scores ranging from 0–3. While searches with more words had higher maximum possible scores than searches consisting of fewer or only one word, we compared the scores of search results between interfaces, not the scores between individual searches. To refine the relevancy scoring further, we added a proximity category to give additional value for searches with natural phrases that correctly appeared in a search result, with closer proximity resulting in a higher score. The proximity category also accommodated the possibility of multiple natural phrases in a single search query.

The authority scale gave a score of 3 to items published by a scholarly press, government organization, or research institute; a 2 to items published or produced for a general audience such as popular, journalistic, or trade publications; a 1 for self-published materials; and a 0 if no publisher was identified.

Finally, we added a penalty that subtracted points for items which appeared more than once in a results list, a natural phrase becoming another natural phrase (e.g., "secondary education" becomes "post-secondary education"), book reviews, and/or Cliff Notes or study guide type materials.

Testing the Rubric and Inter-Rater Reliability

As we created our rubric, we used one day of search queries pulled from Goshen College's Google Analytics data from WorldCat Discovery, in order to evaluate search results with the developing rubric. We chose October 12, 2022 for this data as a random weekday in the middle of a semester. We selected 10 search queries that we believed represented the types of searches entered by users, such as searches for names, multiple phrases, and proper nouns. (See appendix B for the list of terms.) Although the search queries were originally used in October 2022, we did not collect the search results that were the basis of the research until 2023. Two researchers used the rubric to evaluate the first five results from each search tool (discovery layer, aggregator database, and Google Scholar) that resulted from reusing the search queries. The first five results were used for the study in order to mimic typical user behavior. With each set of evaluations, we further refined the rubric.

In order to evaluate the number of search queries used in this study, we needed to divide the work among multiple people. And to ensure that each person applied the rubric similarly, we needed to confirm inter-rater reliability among our research group. To accomplish this, each of our four members evaluated 105 search results, formatted as citations (reusing seven search queries from

our original list) and then we compared the scores. We used the statistical measure coefficient of variation (CV) for comparison purposes. Smaller coefficients of variation indicate consistency of performance.[49] All four members' scores when compared had a CV of 2.02%. We also met regularly via web conferencing software and sent questions to the group via email to iron out some of the specifics of the rubrics.

Once we were confident that we had achieved a sufficient degree of inter-rater reliability, we moved on to the actual body of search queries, using one week of search terms for Goshen College's WorldCat Discovery instance, obtained from Google Analytics. We chose October 23–29, 2022, as a representative sample of a mid-semester week where research projects would be happening. In that week, 616 unique queries were used in WorldCat Discovery (47 were used more than once).

Of those 616 queries, 302 search strings (49%) were determined to be known items where the item appears in the search results of any of the three databases or is a verifiable citation. Our research for this study focused only on keyword searching rather than known item searching. While libraries do need a tool that can do known item searching, whether that be in a library catalog or discovery layer product, our purpose in this study was to focus on users who were not looking for a particular item. Thus, we eliminated those 302 known-item searches for the purposes of this study. We also eliminated 116 queries for the following reasons:

- An author who does not write scholarly material (6 terms)
- Problematic spelling errors (29 terms)
- Search for library resources or locations (4 terms)
- Use of an index not available in all databases (27 terms)
- Search produces fewer than five results in one of the three search tools (39 terms)
- Duplicates another query but with capitalization (11 terms)

This left us with 198 search queries that were used for this study. (See appendix C for the full list of queries used.)

For each query, we entered the search string into Academic Search Complete, Google Scholar, and WorldCat Discovery and extracted citations for the first five results using the products' automated tools. The citations were formatted using the Chicago Manual of Style's notes and bibliography format because it includes authors' first names, which were evaluated as part of the rubric. We also documented when the search terms appeared onscreen as excerpts from the citations or as subject terms.

Goshen College has configured Academic Search Complete to use SmartText Searching. EBSCO describes this feature as "a natural language search strategy that allows you to enter as much text as you like for a search—a phrase, sentence, paragraph, or even whole pages." When a user enters a search string that results in zero search results, SmartText Searching adds relevancy weight to words and phrases, as well as Boolean ORs to the search and runs it again. This happened for 30 of our queries. Google Scholar was not configured in any way for this study. All Google Scholar searches were run on the same computer browser. Several search queries were spot-checked on a separate computer and returned the same results. WorldCat Discovery has been configured by Goshen College to sort by Best Match and show items held by Libraries Worldwide.

Once the citations were obtained, our group evaluated the citations based on the rubric with the citation search source hidden. One researcher evaluated 44.44% (1320 citations), another evaluated 25.25% (750 citations), and the other two evaluated 15.15% (450 citations) each.

RESULTS

Out of a maximum rubric score of 12 for any given search result, citations scored on average 6.5 from WorldCat Discovery, 7.4 from Google Scholar, and 8.0 from Academic Search Complete.

While the average scores provide an overview of the search tools' performance, insights can also be gained by comparing the average scores for each component of the rubric (see table 2). From these component scores, it is evident that Academic Search Complete is heavily prioritizing the newest materials, while Google Scholar has the most relevant items with the exact search terms that researchers are using. Both Academic Search Complete and Google Scholar are fairly close in authority rankings, with WorldCat Discovery lagging behind.

Table 2. Rubric Scores

	Average score	Average currency	Average relevancy	Average proximity	Average authority
Academic Search Complete	8.0	2.4	2.0	0.8	2.9
Google Scholar	7.4	1.4	2.2	1.0	2.8
WorldCat Discovery	6.5	1.5	1.8	0.8	2.5

As discussed above, our rubric did not adjust its scoring for currency based on the discipline of the search query. We know that the value that researchers place on currency varies widely across disciplines, with scientific fields focusing more heavily on recent literature and fields in the humanities finding older citations more acceptable. For our scoring, if currency is removed altogether from the ranking, then Academic Search Complete scored on average 5.6, Google Scholar scored on average 6.0, and WorldCat Discovery scored on average 5.1. This matches the relevancy ranking used by the three products, as Academic Search Complete heavily prioritizes the most recent materials, even if they may not be the most relevant.

DISCUSSION

Our study is unique among the literature in that it sought to create a reproducible scoring rubric that could be applied to actual user searches. Six articles in the literature did include some sort of scoring rubric.[50] However, none were entirely reproducible. Three studies in the literature evaluated searches from usage reports.[51] However, their findings were not evaluated against scoring rubrics that could be reproduced. We hope that future research using shared criteria will help the library community communicate with library vendors (and perhaps even Google) about the specific limitations of these tools. We also believed that previous studies, which only used a binary scale to evaluate relevance, left a gap in the literature that this work could fill. Instead, we wanted to use a graded scale similar to the industry standard of cumulative gain (CG).[52]

Our results were consistent with most of the prior literature reviewed above; we found that Google Scholar did a better job of prioritizing relevance and proximity than the other two search tools. While our study focused on keyword searching, future research should consider known item searching with the same, or similar, tools, as Singley found that WorldCat Local, the precursor to

WorldCat Discovery, outperformed Google Scholar with known item searching.[53] Our results for authority found that EBSCO's Academic Search Complete narrowly outperformed Google Scholar, which is consistent with the work of Asher et al. and Georgas regarding quality of results and EBSCO Discovery Service.[54] While EBSCO Discovery Service's central index is larger than Academic Search Complete's database, and ranking quality of results is a broader category than authority itself, this regardless fell in line with the existing research. Our results are also consistent with research reviewed above which found that aggregator databases tend to prioritize more recent results than Google Scholar and library discovery layers. WorldCat Discovery emphasized recent results slightly less than Google Scholar did, we surprisingly discovered. However, the research that evaluated currency did not include WorldCat Discovery (nor WorldCat Local), but rather other discovery layer tools.

Our goal for this study was to compare the relevancy of results from WorldCat Discovery with other products aimed at novice researchers. We chose EBSCO's Academic Search Complete because Indiana provides all libraries with access via the State Library. We chose Google Scholar because it is free and used by many students and faculty for academic searching. This study deliberately avoided assessment of typical user behavior by not making assumptions about the search terms used. Experienced librarians can of course make guesses about searchers' intentions, but without actually interacting with users, it is impossible to make conclusive judgments about their intentions. We avoided relying on assumptions or subjective measures, and instead evaluated the search results themselves, in a process that paralleled how a typical user would evaluate search results. We likewise avoided direct comparison and contrasting of the search tools in terms of filters, facets, indexing, metadata, or full text availability.

CONCLUSION

Our study asked whether a discovery layer was the best tool for finding relevant results and answered that question by comparing a typical discovery layer with other commonly chosen search tools, using keyword search queries collected from actual users. Ultimately, we found that Academic Search Complete provided the most current and authoritative sources, while Google Scholar provided the most relevant sources, and WorldCat Discovery was the weakest tool for current, relevant, and authoritative results.

Discovery layers are popular and often the first tool encountered on a library homepage but can frustrate users if they do not return relevant results. Before the advent of discovery layers, in response to demands for a "one-stop shop" or "Google-like" search in a library context, information literacy instruction focused almost exclusively on either the online catalog or using one database or index at a time, usually subject-specific. However, if we know the discovery layer will give the least relevant results, why continue to teach students how to use it and focus on it? There is no simple shortcut to get to the best sources for any topic while always using the same tool. Alternatively, a multidisciplinary database provides relevance but searches across fewer resources, which might be just what beginner researchers need as they are still building their information literacy skills. One facet of information literacy is the ability to determine the type and scope of information that will be effective for a specific information need or research question. Users will retrieve more relevant results when they are able to understand more clearly what will satisfy their question or information need before using a "one size fits all topics" search box.

Currently, many libraries have discovery layer search boxes on their homepages, directing researchers to this product as a one-stop location to meet their research needs. However,

discovery layers may not actually be meeting researcher needs and should be reconsidered. As evidenced by librarian experience and research studies, students tend to use the first box they see on a library website without making an intentional choice.[55] Rather than giving students an "easy button" for research, we need to train our students to intentionally think through their research habits and search patterns.

Additionally, after students graduate, they will no longer have access to the same suite of library products that are available to them while they are enrolled. While state and public libraries offer research support to most people in the US, their budgets are much smaller, and the products tend to be more practical (such as a small engine repair database). Teaching students how to use products that are freely available to them, such as Google Scholar, PubMed, and ERIC, will ensure that they can continue to find reliable information as professionals.

Libraries must be honest about the benefits and drawbacks of discovery tools and weigh that with the costs of maintaining these tools especially in the era of generative artificial intelligence. Are the drawbacks of discovery tools significant enough to warrant a reconsideration of their widespread adoption by academic libraries? Is supporting a tool that does not produce the most current, relevant, or authoritative sources, compared to other existing search products, worth the cost compared to the benefit of providing apparently—although not actually—one-stop searching?

APPENDIX A: RUBRIC

Items were scored positively on a scale of 0 to 3 or penalized on a scale of -1 or -0.5 using the following guidelines.

Currency

Score	Guidelines
0	More than 50 years old (published before 1974)
1	More than 10 years old (published before 2014)
2	Published within the past 6–10 years (published between 2014 and 2019)
3	Published within the last 5 years (published between 2020 and 2024)
Notes	If serial publications are still being published, they will receive a 3; otherwise, they will be scored based on when they ceased publication

Relevancy

Score	Guidelines
0	0 of the search terms appear in the citation or the result is not in the language searched. If the search terms were an author, no terms found in the citation.
1	1 of the search terms or synonyms for search terms appear in the citation. If the search terms were an author, first name only found in the citation.
2	2 of the search terms or synonyms for search terms appear in the citation. If the search terms were an author, last name only found in the citation.
3	3 or more of the search terms or synonyms for search terms appear in the citation. If the search terms were an author, full name found in the citation.
Notes	• If the search terms do not appear in the citation or author, but DO appear on screen in the brief results, score the onscreen search terms using the criteria above. • Plural and singular versions of terms should be counted as the same for scoring purposes. • Search terms that are articles, Boolean operators, or two-letter prepositions are not included in this score: a, an, the, and, or, not, of, on, to, in, etc. • Author terms must either both be found in the author or title field to receive a 3.

Proximity

Score	Guidelines
0	Natural phrases in the title or author are not 1 word proximity or are not in the language searched.
1	At least one of the natural phrases in the title or author is 2 word proximity.
2	At least one of the natural phrases in the title or author is 1 word proximity.
3	Two or more of the natural phrases in the title or author are 1 word proximity.
Notes	• Any variant forms of the terms should be counted as the same for scoring purposes (such as plural, possessive, past tense, etc.). • 2 word proximity = one word in between the words of the natural phrase: e.g., patients and falls vs patient falls. • 1 word proximity = directly beside each other . • Natural phrase = a group of words that go together, e.g., John Fitzgerald Kennedy, assassination attempt, United States, boy scout. Does not include synonyms.

Score	Guidelines
	• Use the simplest natural phrases from the search term, e.g., teaching of American literature, natural phrase = American literature. • Natural phrases must be in the original order. • Proximity scores are given for names.

Authority

Score	Guidelines
0	No publisher identified.
1	Published or produced with no accountability (self-published or vanity press).
2	Published or produced by a commercial or professional organization for a general audience (popular, journalistic, or trade publications). Includes religious publishers not affiliated with higher education.
3	Published or produced by a scholarly press, government organization, or research institute. Includes dissertations and non-English language items.

Penalty

Score	Guidelines
-1	Result repeated from earlier in the results list
-1	Natural phrase has become another natural phrase, e.g., secondary education is now post-secondary education
-1	Book review
-.5	Cliff Notes or study guide materials

APPENDIX B: SEARCH TERMS USED IN RUBRIC CREATION

an american constitution
anabaptist
assistance animals history
bhutan
birds power lines electrocution
Christina Hunger
contemplative education
patient falls nurse staffing
Susan Cain
Teaching of American literature

APPENDIX C: SEARCH TERMS USED IN FULL STUDY

1984

21st century skills

21st century skills esl

Abdulla, Munawwar

abnormal psychology

abortion

Abortion in the United States

abortion laws and ethical considerations

Abortion laws in the United States

abortion laws united states

abortion laws us

abortion rights

abortion rights united states

academic advising in small groups

adolescent development English

adolescent development esl

adolescent loneliness

adolescents loneliness

adolescents loneliness clubs

adolescents loneliness virtual reality

Africa

africa children book

amazon success

amish culture

Angola

animal language

anti-asian violence in detroit

argentina dirty war

art nouveau

asian american identity

assessments for ADHD

au="Chung, Sue Fawn"

Australia

Australia childrens book

Australia childrens books

Bachelor parties

Bachelor party

ben and jerrys

ben and jerrys unilever

boarding schools

Bodies and boundaries

book of Luke

book of Revelation

both sides of gun control

brian blount

cabo verde

Cameroon

Carreira Antonio

chicano history

children's books

childrens books from tanzania

china town in chicago

chinese american death

chinese american exhumation

chinese cuban

chinese exhumation

chinese prostitution

chinese railroad worker writings

Chinese Returning dead home

Christianity and El Salvador

citizen

classism among asian americans

community knowledge middle school science

congo

consequences of Russian revolution

Consequences of the Russian revolution

continuous deep sedation

contraception and gender equality

correct idioms

crime and gender in Europe

daniel

Darren byler

Deaf communication

death

death psychology

depression

depression and anxiety

developmentally appropriate english class

dozier school

dozier school for boys

Dr. Patricia Islas Salinas

drawdown

ecclesiology mennonite

english

english 21st century skills

ezekiel

ezekiel bible

ezekiel commentary

family planning climate change

family planning gender equality

fiction

formative assessment AND middle school science

formula feeding effects on infants

Garen, A.,

Gender Asia

gender equality climate change population growth

Gender Middle East

Genesis

Georgia okeefe

Getachew

global citizenship college

global families

gospel of Luke

gun control and crime

gun reduce gun violence

gun violence

gun violence in the united states

gun violence statistics

Hadith

hmong

how mimicry can hurt interactions

ibuprofen side effects

identity development in adolescence

identity education in adolescence

idioms

indian immigration to the midwest

Interpretation

interpretation daniel

interpretation of daniel

interpreting

is abortion a religious issue

Isaiah

Jane Adams

Japanese resettlement

jesus and the lawyer

juvenile justice sentencing

Kaebnick

Kenya

kenyan

kenyan childrens books

korean american christian

lgbt psychology

literary devices Revelation

Louis f hartman

macro and micro scale techniques

madagascar

marching band

me too movement

mediation

mediation family

Mediation in political

Memoir

Memoir graphic

mennonite theology

mental illness

micro and macro

Middle school development english

Middle school development esl

Middle school developmental english

minnesota starvation experiment

moral development in adolescence

music and influences

Mystery

new zealand

Nigeria Childrens Book

Oceania

Oppressing

pancreatic cancer

People and Prayer

plate tectonics

population growth and climate change

psychology

psychology and death

psychology death

psychology of depression

regeneration

religion and politics the united states abortion

religion and women's rights

Religion Gay rights

revelation

Revelation 12

revelation commentary

revelation literary devices

rhetoric in the book of Revelation

Rhetorical devices Revelation

section 8 housing

section 8 legislation

sex education

sex education benefits

sex education economic opportunity

should marching band be considered a sport

social justice in the esl classroom

social work liberal conservative

south asia feminism

studies of assessments for ADHD

tafsir

tanzanian music

tesol

the book of daniel

ti:ezekiel commentary

ti:gun violence

vulgate

walkable cities

Wanda

what is important about a pitch

William Cavanaugh

wombat

Xylophone

ENDNOTES

[1] "Perceptions of Libraries and Information Resources" (Dublin, Ohio: OCLC Online Computer Library Center, Inc., 2005), https://www.oclc.org/content/dam/oclc/reports/pdfs/Percept_all.pdf.

[2] Burton Callicott and Debbie Vaughn, "Google Scholar vs. Library Scholar: Testing the Performance of Schoogle," *Internet Reference Services Quarterly* 10, no. 3–4 (2005): 71–88, https://doi.org/10.1300/J136v10n03_08; Susan Gardner and Susanna Eng, "Gaga over Google? Scholar in the Social Sciences," *Library Hi Tech News* 22, no. 8 (January 1, 2005): 42–45, https://doi.org/10.1108/07419050510633952; D. Yvonne Jones, "Biology Article Retrieval from Various Databases: Making Good Choices with Limited Resources," *Issues in Science and Technology Librarianship* 44 (2005), http://www.istl.org/05-fall/refereed.html.

[3] Jenny S. Bossaller and Heather Moulaison Sandy, "Documenting the Conversation: A Systematic Review of Library Discovery Layers," *College & Research Libraries*, July 14, 2017, https://doi.org/10.5860/crl.78.5.602; Rachel Cooke and Rebecca Donlan, "Thinking inside the Box: Comparing Federated Search Results from Google Scholar, Live Search Academic, and Central Search," *Journal of Library Administration* 46, no. 3–4 (March 6, 2008): 31–42, https://doi.org/10.1300/J111v46n03_03; Kate B. Moore and Courtney Greene, "Choosing Discovery: A Literature Review on the Selection and Evaluation of Discovery Layers," *Journal of Web Librarianship* 6, no. 3 (July 2012): 145–63, https://doi.org/10.1080/19322909.2012.689602.

[4] Elizabeth Ketterman and Megan E. Inman, "Discovery Tool vs. PubMed: A Health Sciences Literature Comparison Analysis," *Journal of Electronic Resources in Medical Libraries* 11, no. 3 (July 1, 2014): 115–23, https://doi.org/10.1080/15424065.2014.938999; Michael Levine-Clark and Joseph Kraus, "Finding Chemistry Information Using Google Scholar: A Comparison with Chemical Abstracts Service," *Science & Technology Libraries* 27, no. 4 (August 20, 2007): 3–17, https://doi.org/10.1300/J122v27n04_02; Mary Shultz, "Comparing Test Searches in PubMed and Google Scholar," *JMLA: Journal of the Medical Library Association* 95, no. 4 (October 2007): 442–45, https://doi.org/10.3163/1536-5050.95.4.442; Simona Ştirbu et al., "The Utility of Google Scholar When Searching Geographical Literature: Comparison With Three Commercial Bibliographic Databases," *The Journal of Academic Librarianship* 41, no. 3 (May 1, 2015): 322–29, https://doi.org/10.1016/j.acalib.2015.02.013.

[5] Arkadiusz Pulikowski and Anna Matysek, "Searching for LIS Scholarly Publications: A Comparison of Search Results from Google, Google Scholar, EDS, and LISA," *The Journal of Academic Librarianship* 47, no. 5 (September 2021): 102417, https://doi.org/10.1016/j.acalib.2021.102417.

[6] Karen Ciccone and John Vickery, "Summon, EBSCO Discovery Service, and Google Scholar: A Comparison of Search Performance Using User Queries," *Evidence Based Library and Information Practice* 10, no. 1 (March 15, 2015): 34–49, https://doi.org/10.18438/B86G6Q; David Earl Noe, "Replicating Top Users' Searches in Summon and Google Scholar," in *Planning and Implementing Resource Discovery Tools in Academic Libraries* (ISR, 2012), 225–49, https://doi.org/10.4018/978-1-4666-1821-3.ch013.

[7] Sarah P. C. Dahlen and Kathlene Hanson, "Preference vs. Authority: A Comparison of Student Searching in a Subject-Specific Indexing and Abstracting Database and a Customized Discovery

Layer," *College & Research Libraries* 78, no. 7 (October 30, 2017): 878–97, https://doi.org/10.5860/crl.78.7.878.

[8] Boram Lee and EunKyung Chung, "An Analysis of Web-Scale Discovery Services from the Perspective of User's Relevance Judgment," *The Journal of Academic Librarianship* 42, no. 5 (September 1, 2016): 529–34, https://doi.org/10.1016/j.acalib.2016.06.016; Tao Zhang, "User-Centered Evaluation of a Discovery Layer System with Google Scholar," in *Design, User Experience, and Usability. Web, Mobile, and Product Design*, ed. Aaron Marcus, vol. 8015, Lecture Notes in Computer Science (Springer Berlin Heidelberg, 2013), 313–22, https://doi.org/10.1007/978-3-642-39253-5_34

[9] Callicott and Vaughn, "Google Scholar vs. Library Scholar"; Rosie Hanneke and Kelly K. O'Brien, "Comparison of Three Web-Scale Discovery Services for Health Sciences Research," *Journal of the Medical Library Association : JMLA* 104, no. 2 (April 2016): 109–17, https://doi.org/10.3163/1536-5050.104.2.004; Pulikowski and Matysek, "Searching for LIS Scholarly Publications."

[10] Ciccone and Vickery, "Summon, EBSCO Discovery Service, and Google Scholar."

[11] William H. Walters, "Google Scholar Search Performance: Comparative Recall and Precision," p*ortal: Libraries and the Academy* 9 (January 1, 2009): 5–24, https://doi.org/10.1353/pla.0.0034; William H. Walters, "Comparative Recall and Precision of Simple and Expert Searches in Google Scholar and Eight Other Databases," *portal: Libraries and the Academy* 11, no. 4 (2011): 971–1006, https://doi.org/10.1353/pla.2011.0042.

[12] Walters, "Google Scholar Search Performance."

[13] D. Yvonne Jones, "Biology Article Retrieval from Various Databases: Making Good Choices with Limited Resources," *Issues in Science and Technology Librarianship* 44 (2005), http://www.istl.org/05-fall/refereed.html.

[14] Blake L. Galbreath, Corey Michael Johnson, and Alexander N. Merrill, "A Framework for Measuring Relevancy in Discovery Environments," *Information Technology and Libraries* 40, no. 2 (2021): 1–17, https://doi.org/10.6017/ital.v40i2.12835; Elizabeth Namei and Christal A. Young, "Measuring Our Relevancy: Comparing Results in a Web-Scale Discovery Tool, Google & Google Scholar," 2015, https://alair.ala.org/bitstream/handle/11213/17887/MeasuringOurRelevancy.pdf?sequence=1; Alexa L. Pearce, "Discovery and the Disciplines: An Inquiry into the Role of Subject Databases through Citation Analysis," *College & Research Libraries* 80, no. 2 (March 1, 2019): 195–214, https://doi.org/10.5860/crl.80.2.195; Emily Singley, "Discovery Systems – Testing Known Item Searching," *Usable Libraries* (blog), March 18, 2014, https://emilysingley.net/discovery-systems-testing-known-item-searching.

[15] Jan Brophy and David Bawden, "Is Google Enough? Comparison of an Internet Search Engine with Academic Library Resources," *Aslib Proceedings* 57, no. 6 (January 1, 2005): 498–512, https://doi.org/10.1108/00012530510634235; Jared L. Howland et al., "How Scholarly Is Google Scholar? A Comparison to Library Databases," *College & Research Libraries* 70, no. 3 (2009), https://doi.org/10.5860/0700227.

[16] Andrew D. Asher, Lynda M. Duke, and Suzanne Wilson, "Paths of Discovery: Comparing the Search Effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and Conventional Library Resources," *College & Research Libraries* 74, no. 5 (2013), https://doi.org/10.5860/crl-374; Dahlen and Hanson, "Preference vs. Authority"; Helen Georgas, "Google vs. the Library (Part III): Assessing the Quality of Sources Found by Undergraduates," p*ortal: Libraries and the Academy* 15, no. 1 (2015): 133–61, https://doi.org/10.1353/pla.2015.0012.

[17] Gardner and Eng, "Gaga over Google?; Cooke and Donlan, "Thinking inside the Box."

[18] Gardner and Eng, "Gaga over Google?"

[19] Cooke and Donlan, "Thinking inside the Box."

[20] Ciccone and Vickery, "Summon, EBSCO Discovery Service, and Google Scholar"; Gardner and Eng, "Gaga over Google?"; Georgas, "Google vs. the Library (Part III)"; Namei and Young, "Measuring Our Relevancy"; Noe, "Replicating Top Users' Searches in Summon and Google Scholar"; Pulikowski and Matysek, "Searching for LIS Scholarly Publications"; Walters, "Google Scholar Search Performance"; Walters, "Comparative Recall and Precision of Simple and Expert Searches in Google Scholar and Eight Other Databases"; Zhang, "User-Centered Evaluation of a Discovery Layer System with Google Scholar."

[21] Georgas, "Google vs. the Library (Part III)"; Namei and Young, "Measuring Our Relevancy"; Pulikowski and Matysek, "Searching for LIS Scholarly Publications."

[22] Asher, Duke, and Wilson, "Paths of Discovery"; Georgas, "Google vs. the Library (Part III)."

[23] Dahlen and Hanson, "Preference vs. Authority"; Gardner and Eng, "Gaga over Google?"; Georgas, "Google vs. the Library (Part III)"; Noe, "Replicating Top Users' Searches in Summon and Google Scholar"; Pearce, "Discovery and the Disciplines"; Shultz, "Comparing Test Searches in PubMed and Google Scholar."

[24] Jones, "Biology Article Retrieval from Various Databases"; Levine-Clark and Kraus, "Finding Chemistry Information Using Google Scholar"; Pearce, "Discovery and the Disciplines"; Știrbu et al., "The Utility of Google Scholar When Searching Geographical Literature."

[25] Richard Wisneski, "I Can't Get No Satis-Searching: Reassessing Discovery Layers in Academic Libraries," *Journal of Web Librarianship* 18, no. 1 (2024): 1–14, https://doi.org/10.1080/19322909.2024.2326687; Casey D. Hoeve and Christina Geuther, "The Silent Battle on the Budget: The Effect of Centralized Indexing on Collection Analysis in Primo and EBSCO," *Collection Management* 46, no. 1 (December 5, 2020): 3–13, https://doi.org/10.1080/01462679.2020.1790458; Xiaotian Chen, "EBSCO Collections' Discoverability Rate by Ex Libris' Central Discovery Index (CDI)," *Collection Management* 48, no. 2 (April 3, 2023): 84–96, https://doi.org/10.1080/01462679.2022.2081277.

[26] Kelly Achenbach et al., "Defining Discovery: Is Google Scholar a Discovery Platform? An Essay on the Need for a New Approach to Scholarly Discovery," *Open Research Europe* 2 (June 7, 2022): 28, https://doi.org/10.12688/openreseurope.14318.2.

[27] Cristífol Rovira, Frederic Guerrero-Solé, and Lluís Codina, "Received Citations as a Main SEO Factor of Google Scholar Results Ranking," *Profesional de la información* 27, no. 3 (June 18,

2018): 559–69, https://doi.org/10.3145/epi.2018.may.09; Cristòfol Rovira, Lluís Codina, and Carlos Lopezosa, "Language Bias in the Google Scholar Ranking Algorithm," *Future Internet* 13, no. 2 (February 2021): 31, https://doi.org/10.3390/fi13020031.

[28] Paul T. von Hippel and Stuart Buck, "Improve Academic Search Engines to Reduce Scholars' Biases," *Nature Human Behaviour* 7, no. 2 (February 2023): 157–58, https://doi.org/10.1038/s41562-022-01518-0.

[29] N. S. Redkina, "Open Scientific Content Search Tools for Research and Education," *Professional Education in the Modern World* 13, no. 4 (2023): 648–60, https://doi.org/10.20913/2618-7515-2023-4-6; S. Khalid et al., "On the Current State of Scholarly Retrieval Systems," *Engineering, Technology & Applied Science Research* 9, no. 1 (February 16, 2019): 3863–70, https://doi.org/10.48084/etasr.2448.

[30] Khalid et al., "On the Current State of Scholarly Retrieval Systems."

[31] Wisneski, "I Can't Get No Satis-Searching."

[32] Asher, Duke, and Wilson, "Paths of Discovery," 472; Callicott and Vaughn, "Google Scholar vs. Library Scholar," 87; Cooke and Donlan, "Thinking inside the Box," 2; Howland et al., "How Scholarly Is Google Scholar?", 228; Pearce, "Discovery and the Disciplines," 196.

[33] Daniel W. Eller, "Transparency and the Future of Semantic Searching in Academic Libraries," ed. Bonnie Lawlor, *Information Services & Use* 42, no. 3–4 (December 16, 2022): 453–61, https://doi.org/10.3233/ISU-220175.

[34] Lucy Kiester and Clara Turp, "Artificial Intelligence behind the Scenes: PubMed's Best Match Algorithm," *Journal of the Medical Library Association : JMLA* 110, no. 1 (2022): 15–22, https://doi.org/10.5195/jmla.2022.1236.

[35] Jayne Dickson, "WorldCat Discovery Services Beta Is Launched," California Digital Library, April 30, 2014, https://cdlib.org/cdlinfo/2014/04/30/worldcat-discovery-services-beta-is-launched/.

[36] "WorldCat Discovery Resources," OCLC, 2024, https://www.oclc.org/en/worldcat-discovery/resources.html.

[37] Jay Holloway, "WorldCat Discovery Product Insights" (May 15, 2024).

[38] "Relevancy and Scoping," OCLC Support, January 12, 2024, https://help.oclc.org/Librarian_Toolbox/OCLC_Service_Configuration/WorldCat_Discovery_and_WorldCat_Local/020Relevancy_and_Scoping.

[39] "Relevancy and Scoping."

[40] "Relevancy and Scoping."

[41] "Academic Search," Library of Congress Catalog, accessed June 28, 2024, https://lccn.loc.gov/sn97001287.

[42] "How Is Relevance Ranking Determined in EBSCO Discovery Service (EDS)?", EBSCO Connect, January 13, 2020, https://connect.ebsco.com/s/article/How-is-relevance-ranking-determined-in-EBSCO-Discovery-Service-EDS?language=en_US.

[43] "About Google Scholar," Google, accessed May 10, 2024, https://scholar.google.com/intl/en/scholar/about.html.

[44] Taryn Marks and Avery Le, "Increasing Article Findability Online: The Four Cs of Search Engine Optimization," *Law Library Journal* 109, no. 1 (Winter 2017): 83–99, https://doi.org/10.2139/ssrn.3204550.

[45] "Google Scholar Help," Google, accessed April 26, 2024, https://scholar.google.com/intl/en%20/scholar/inclusion.html.

[46] Joran Beel and Bela Gipp, "Google Scholar's Ranking Algorithm: The Impact of Citation Counts (An Empirical Study)," in *2009 Third International Conference on Research Challenges in Information Science* (2009 Third International Conference on Research Challenges in Information Science (RCIS), Fez, Morocco: IEEE, 2009), 439–46, https://doi.org/10.1109/RCIS.2009.5089308.

[47] Beel and Gipp, 7.

[48] Howland et al., "How Scholarly Is Google Scholar?"; Georgas, "Google vs. the Library (Part III)."

[49] Seemon Thomas, *Basic Statistics* (Alpha Science International Ltd., 2014), https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=5190782.

[50] Asher, Duke, and Wilson, "Paths of Discovery"; Brophy and Bawden, "Is Google Enough?"; Cooke and Donlan, "Thinking inside the Box"; Dahlen and Hanson, "Preference vs. Authority"; Gardner and Eng, "Gaga over Google?"; Georgas, "Google vs. the Library (Part III)."

[51] Ciccone and Vickery, "Summon, EBSCO Discovery Service, and Google Scholar"; Namei and Young, "Measuring Our Relevancy"; Noe, "Replicating Top Users' Searches in Summon and Google Scholar."

[52] Nate Day, "Choosing Your Search Relevance Evaluation Metric," *OpenSource Connections* (blog), February 28, 2020, https://opensourceconnections.com/blog/2020/02/28/choosing-your-search-relevance-metric/.

[53] Singley, "Discovery Systems – Testing Known Item Searching."

[54] Asher, Duke, and Wilson, "Paths of Discovery"; Georgas, "Google vs. the Library (Part III)."

[55] Suzanna Conrad and Christy Stevens, "'Am I on the Library Website?': A LibGuides Usability Study," *Information Technology and Libraries* 38, no. 3 (September 15, 2019): 50, https://doi.org/10.6017/ital.v38i3.10977.