The Boolean Is Dead, Long Live the Boolean! Natural Language versus Boolean Searching in Introductory Undergraduate Instruction

M. Sara Lowe, Bronwen K. Maxson, Sean M. Stone, Willie Miller, Eric Snajdr, and Kathleen Hanna

Boolean logic can be a difficult concept for first-year, introductory students to grasp. This paper compares the results of Boolean and natural language searching across several databases with searches created from student research questions. Performance differences between databases varied. Overall, natural search language is at least as good as Boolean searching. With evidence that students struggle to grasp Boolean searching, and may not use it even after instruction, it could be left out of first-year instruction, freeing up valuable class time to focus on concepts such as question development and source evaluation. As the Framework for Information Literacy does not specifically address Boolean operators, the authors suggest it should have less prominence in first-year Information Literacy instruction.

Introduction

Conventional wisdom considers knowledge of Boolean logic a basic information retrieval Information Literacy (IL) skill. Librarians and other information professionals are taught the value of Boolean searching (referred to throughout this article interchangeably as “Boolean”) in professional education, and it is seen in instruction, reference interactions, and database interfaces. However, the concept can be difficult for first-year (introductory) students to grasp, and it can take multiple sessions before a student demonstrates effective use of Boolean logic.

A student's ability to use Boolean operators is a performance indicator within the Association of College and Research Libraries (ACRL) Information Literacy Compe-

*M. Sara Lowe is Educational Development Librarian in the University Library at Indiana University-Purdue University Indianapolis; e-mail: mlove@iupui.edu. Bronwen K. Maxson is Romance Languages Librarian in Norlin Library at University of Colorado Boulder; e-mail: bronwen.maxson@colorado.edu. Sean M. Stone is the Librarian at the Indiana University School of Dentistry; e-mail: smstone@iu.edu. Willie Miller is Informatics & Journalism Librarian in the University Library at Indiana University-Purdue University Indianapolis; e-mail: wmmiller@iupui.edu. Eric Snajdr is Science Librarian in the University Library at Indiana University-Purdue University Indianapolis; e-mail: esnajdr@iupui.edu. Kathleen Hanna is Physical Education/Tourism Management Librarian in the University Library at Indiana University-Purdue University Indianapolis; e-mail: kgreatba@iupui.edu. ©2018 M. Sara Lowe, Bronwen K. Maxson, Sean M. Stone, Willie Miller, Eric Snajdr, and Kathleen Hanna, Attribution-NonCommercial (http://creativecommons.org/licenses/by-nc/4.0/) CC BY-NC.

doi:10.5860/crl.79.4.517
tency Standards for Higher Education. These standards, though rescinded in 2016, influenced IL education across the United States and beyond for more than fifteen years. In studies of first-year students’ IL skills, demonstrated knowledge of Boolean logic is frequently evaluated as a determinant of information retrieval proficiency. This has led to many librarians teaching Boolean logic in one-shot instruction sessions, first-year IL modules and tutorials, and reference interactions. Anecdotally, the authors have worked at multiple institutions and taught Boolean regularly, generally introducing the concept at the first-year (introductory) level, and building on that in upper-level classes. Interestingly, while the ACRL Standards specifically mention Boolean, the new Framework for Information Literacy for Higher Education only refers to searching (controlled vocabulary, keywords, natural language).

For students, it is not always clear that using Boolean logic is better than a natural language or phrase search in the style of Google. Anecdotally, librarians have seen students’ natural language searches yield relevant results explaining why students often find Boolean searching superfluous. Librarians and faculty could use time dedicated to teaching Boolean logic to teach other IL concepts (such as question development and source evaluation) in courses or disciplines in which Boolean searching is not essential. This is particularly the case in introductory courses. Moreover, the instruction of technical Boolean logic is out of sync with the Framework and its less mechanical, more conceptual approach to IL.

Natural language searching has various definitions. For the purposes of this article, we define natural language as searching in phrases or sentences instead of a structured search query using operators and/or punctuation. Since their creation in the 1990s, natural language search algorithms have improved dramatically. However, there have not been any recent studies evaluating the efficacy of natural language searching compared to Boolean searching, and many of the older studies were from an information retrieval, rather than a pedagogical, perspective.

If using Boolean logic correctly is challenging for students, is poorly used, and has no clear advantage in using it to retrieve relevant results, why are we trying to teach it to first-year, introductory students? This study sought to answer the question: Was there any advantage, based on the relevance of search results, to teaching Boolean to first-year students? This article investigates the efficacy of retrieval of Boolean searching compared to Google-style natural language searching for simple research questions of the kind used in research projects for many introductory undergraduate courses.

Literature Review
There are three areas most relevant to the current study: 1) how students search for information; 2) the effectiveness and limitations of Boolean and natural language searching; and 3) how librarians teach Boolean in information literacy instruction.

There is extensive research on how students search for and find information, primarily the “principle of least effort” noted famously in library literature by Mann. Convenience is a major factor; multiple studies found that students prefer to search for information in the easiest possible way to complete their research quickly. This generally translates into students spending little time evaluating search results, often not venturing past the first article and rarely moving beyond the first page of results. Students are effectively turning over the evaluation process to the search algorithm.

Specifically regarding Boolean, Dempsey and Valenti note that students in an introductory English Composition course used odd combinations of Boolean (for instance, “NOT kids AND kids”, accidental use of OR, confusing AND and OR) demonstrating unfamiliarity with how to effectively use the connectors. Boolean use in OPACs, library databases, and the web has been copiously studied, most concluding that users
have trouble understanding how to use Boolean to get relevant results and often use Boolean incorrectly. Multiple studies have found only a small percentage of students use Boolean in their searches, and that is usually limited to the AND operator.

In addition, when breaking their research question or thesis into keywords (Boolean-type searches), students tend to search a limited number of terms, usually around two. But, when they can pose their information need as a question (natural language, phrase searching), they use more terms, which may actually help improve retrieval of relevant results and, most important, increase students’ self-efficacy in the search process. Indeed, as Taylor noted in 1962, asking users to translate a complex question into keywords is an oversimplification of their information need.

Boolean logic for information retrieval has been used since the 1960s, and is just one of many retrieval models (such as vector space or fuzzy set). Decades of information retrieval literature note the limitations of Boolean searching. Of particular importance to the current study is that Boolean searching can be difficult for novice users and often requires a trained intermediary (such as a librarian) to help craft an effective search.

The literature indicates that even people trained in the mathematical concept of Boolean logic have trouble applying it to information retrieval. The 1990s heralded a new retrieval model, commonly called natural language, because users do not need to enter searches as Boolean statements. Multiple studies in the ’90s found a slight edge to Boolean over natural language in retrieval of relevant results but not to the degree that one would be preferable over the other. A 1995 article evaluating Boolean versus natural language (phrase) searching in full-text online medical textbooks found there was no statistically significant difference in recall or precision between the types of searches. Conversely, Turtle, in 1994, found natural language queries were more effective (for instance, they retrieved more relevant results) than Boolean, even for expert searchers comfortable with Boolean.

Research has shown that systems can be developed to produce more accurate natural language searches.

Beyond understanding student use of search strategies, it is also important to understand how librarians are teaching Boolean. There is surprisingly little in the literature about exactly how Boolean is taught to students, at what student level, and to what effect. The literature offers more from a theoretical perspective.

Although discussing Boolean in contrast to a discovery system, which is outside the scope of the present article, Cmor and Li in 2012 noted that not having to teach Boolean would free up instruction time that could be used to delve more deeply into evaluating and engaging with sources rather than dealing with the mechanics of a search. Similar conclusions were drawn by Buck and Mellinger in 2011 who, in a survey of librarians using Summon, found respondents were spending more time teaching threshold concepts such as the research process, evaluating results, and peer review, and less time on Boolean and database choice. It is interesting that both of these studies predate, but address issues presented in, the Framework. This speaks to a disconnect in the profession between the Standards and the Framework as well as the reasons for, and the efficacy of, teaching Boolean.

In the Dempsey and Valenti study mentioned above, the authors describe their teaching methodology, which involves selecting keywords and evaluating sources but does not specifically include Boolean. The curriculum Burns discusses in a nursing first-year seminar does note that students must use Boolean terminology in their searches but does not address how students used Boolean nor does it provide any assessment of students’ skills. Quarton explains how faculty can develop an assignment so that
students can effectively write a research paper, including developing keywords and search statements using the Boolean operator AND, but with no assessment of whether students learn and retain this knowledge.27

While the present study does not include assessment of student work, there are studies that support our findings that do. Multiple studies examine assessment of library instruction of Boolean search techniques. Lacy and Chen discuss a curriculum for an introductory composition course in which they taught and assessed student use of Boolean terms. Highlighting the complexity of Boolean for novice users, the authors note that, after instruction all students structured a search using Boolean at least once, “even if they eventually reverted to natural syntax (single keywords or phrases)” [emphasis added].28 Novotny studied student learning after an instruction session on searching the OPAC and found limited evidence that students effectively applied Boolean search techniques, with a few using advanced search strategies and the Boolean OR.29 But this was not a longitudinal study, so it is unclear if students continued to use these strategies after the instruction session. Vine bluntly stated: “And then they will leave the class, go home, and... they will revert to the ‘plug-in-the-keyword’ approach.”30

Theoretically, the move from the ACRL Standards for Information Literacy31 to the Framework for Information Literacy for Higher Education32 would tend to indicate either a lessening of the importance of Boolean or, perhaps more accurately, an understanding of broadening the focus to search strategies rather than a narrow emphasis on Boolean. As mentioned, while the ACRL Standards specifically reference Boolean, the Framework only refers to searching (controlled vocabulary, keywords, natural language). In other words, while the Standards presented more of a rote, mechanical model in specifically mentioning Boolean, the Framework allows librarians to take a conceptual approach in teaching search strategies. Multiple authors have addressed the Framework and how it offers a different and promising approach to Information Literacy instruction for librarians, faculty, and students.33 A two-part study by Scott indicates undergraduate students can grasp these concepts.34

Foasberg highlights how the origins of the present study are exactly represented in the Framework: “The Framework’s embrace of constructivist philosophy—which holds that knowledge is constructed and reconstructed through social interactions—makes it less reductive and more inclusive than the Standards’ positivist approach, which assumes that information is objective and measurable.”35 Why do we teach Boolean? Questioning that seemingly fundamental IL skill is, in effect, the function of the Framework.

This research compares the efficacy of Boolean versus natural language searching in light of current database functionality to determine the usefulness of including Boolean instruction, including limiters, search persistence (like going beyond the first page), and variations within and between databases, in introductory undergraduate courses. We sought to answer several questions. Do Boolean searches retrieve more relevant results than natural language searches? Is there any advantage to using filters such as the scholarly article limiter? How accurate are database relevance ranking algorithms (on and beyond the first page) and should librarians be concerned if students do not go beyond the first page?

**Methodology**

We began by doing an initial search of Academic Search Premier, with Boolean and natural language queries based on a sample research topic. For simplicity’s sake, and because most students rely solely on it, we only used the Boolean operator AND. The authors performed each query twice from the basic search screen, first with no limiters and then using the “scholarly (peer reviewed) journals” filter. Four searches
were performed: unfiltered Boolean; filtered Boolean; unfiltered natural language; and filtered natural language. We evaluated the first twenty-five search results (n = 25) for relevance using a rubric (see table 1 for rubric and table 2 for norming question). The rubric was normalized during this test process and led to additional descriptive language for each category including the “or” statements, which widened the inclusion of articles in various categories on the scale of 0 = not relevant to 3 = very relevant.

### TABLE 1
**Article Relevance Evaluation Rubric**

<table>
<thead>
<tr>
<th>Not Relevant (0)</th>
<th>Less Relevant (1)</th>
<th>Relevant (2)</th>
<th>Very Relevant (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 of total concepts represented OR false hits, terms are there but used in different ways (e.g., social work instead of social rejection)</td>
<td>Less than half concepts represented OR concepts are there but not relevant to research question</td>
<td>Majority or all of concepts represented either in title or abstract but when looking at abstract, may be tangential to research question</td>
<td>All concepts represented in title or abstract and abstract is relevant</td>
</tr>
</tbody>
</table>

### TABLE 2
**Norming Research Topic & Search Query**

<table>
<thead>
<tr>
<th>What are the effects of social rejection on lesbians?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boolean query: Social rejection and lesbians</td>
</tr>
<tr>
<td>Natural language query: Effects of social rejection on lesbians</td>
</tr>
</tbody>
</table>

After the rubric was normed, we created three research projects and query sets based on actual student research questions similar to those seen in introductory, undergraduate courses. These were designed to reflect diversity in subject material and database scope and content. From these questions, we created standardized queries as were used in the testing phase. Authors formed three interrater pairs and searched each of the three project queries in two or three of eight databases (see table 3). Searches were completed in 2016 between July 5 and August 3.

Authors chose databases based on the perceived likelihood of an undergraduate encountering and using them to find resources for an introductory research project or paper. In an effort to cover as many disciplines as possible, the authors included: Academic Search Premier (EBSCO); Google Scholar; JSTOR; LexisNexis Academic; ProQuest Central; PubMed; Scopus; and Web of Science. There has been quite a bit of previous research comparing various databases and Google Scholar, often focused on STEM disciplines and primarily comparing Scholar to Web of Science, Scopus, and PubMed. Each database was searched in the same manner as Academic Search Premier (EBSCO) was in the initial test phase, with four searches when possible: unfiltered Boolean; filtered Boolean; unfiltered natural language; and filtered natural language. Filtered searches included limits for scholarly/peer-reviewed articles or the closest equivalent when available: Google Scholar could not be limited in this way; JSTOR was limited to “Articles”; LexisNexis Academic, to “Law Reviews”; ProQuest Central, to “Peer reviewed”; PubMed, to “Journal Article”; Scopus and Web of Science, to “Article.”
The authors captured the first twenty-five search results with screenshots or other export tools for later comparison (n = 25 for all databases and all searches, except for the filtered natural language search in Academic Search Premier, for which only 20 results were returned). We recorded rubric scores for relevance as well as overlap between filtered and unfiltered searches. Interrater pairs met to discuss and normalize scores. During norming, the authors determined reading the full text of all articles was infeasible and, more important, not representative of the way students would quickly scan and evaluate articles. Because of this, rubric scores were based solely on the title and abstract.

**Results**

**Overall Results**

As might be expected, general searches such as these returned a rather high number of overall results. Natural language searches retrieved fewer total results than their Boolean counterparts (for example, unfiltered Boolean versus unfiltered natural language) in all databases except JSTOR and LexisNexis Academic, where the trend was reversed (see table 4).

As mentioned above, this study only evaluated the first 25 results from each search using the rubric. Average rubric scores for all four searches (unfiltered Boolean, filtered Boolean, unfiltered natural language, and filtered natural language) ranged from 1.98 to 2.08 (out of a high of 3), solidly in rubric level 2 = relevant (see figure 1). The unfiltered natural language search query received the highest average score of 2.08, and the filtered Boolean search query received the lowest average score of 1.98. Relevance means the majority of all concepts in the search results, in either the title or the abstract, were present, even if they may be tangential to the research question. In other words, a first-year student will likely find information about a topic using either search method, with or without the filters.

When looking at only the first page of the search results, the range shifted up to 2.03–2.11, with the unfiltered natural language search again having the highest rubric scores (see figure 1). (N for the first page varied by database from n = 10 to n = 50;
The Boolean Is Dead, Long Live the Boolean! 523

see database results section for details.) This indicates the results on the first page are slightly more relevant than the full 25 results. Since research shows students do not usually go beyond the first page of search results, it is encouraging that database algorithms return more-relevant articles first. Results suggest the database’s internal relevance algorithm is slightly outperforming the librarian’s Boolean search. A one-way between subjects ANOVA was conducted, and there was not a significant effect between mean rubric scores at the \( P < .05 \) level.

To get a more granular understanding of the relevance of search results, we divided them into thirds. The top third is interesting because, just as research shows students rarely go past the first page of results, “the farther down the first page a result appears, the less critically it is evaluated.” Were there really relevant results further down the page? There is not a wide difference in thirds—but results do get slightly less relevant further down the page (see figure 2). The top third of results (results 1–8) were the most relevant, with the middle and bottom third dropping slightly. On average, these databases are sending more relevant references to the top of the search results, although a one-way between subjects ANOVA was conducted and there was not a significant effect between any of the thirds at the \( P < .05 \) level.

**Results by Database**

When looking at individual, rather than aggregate, database results, a slightly different story emerges, with some databases clearly outperforming others in returning more relevant results (see figure 3). (Unless otherwise mentioned, as mentioned above, \( n = \)
Overall, Academic Search Premier performed the best among all four searches with an average rubric score of 2.56. Google Scholar was a close second, at 2.50. That puts these two databases close to rubric level 3 = very relevant. On the low end were ProQuest Central at 1.25 and LexisNexis Academic at 1.39, both closer to rubric level 1 = less relevant. In the middle, with rubric scores hovering around rubric level 2 = relevant, were JSTOR (2.34), Web of Science (2.18), Scopus (2.12), and PubMed (1.94).

When looking at consistency, or similarity of rubric scores, between the four searches within a database, Academic Search Premier and Google Scholar also performed the best. ProQuest Central had the widest score range, 1.92 for unfiltered Boolean but only 0.52 for filtered natural language. Because this research focused on first-year or
introductory students, it is interesting to note that, of the typical “first-year” or general subject databases, Academic Search Premier and Google Scholar greatly outperformed ProQuest Central. A one-way between subjects ANOVA was conducted to compare the differences between scores within each database confirms this consistency (or lack of consistency in the case of ProQuest Central), with no significant difference at the $P < .05$ level except in the case of ProQuest.\

Returning to rubric scores from the first page of results, but this time by database, there is some fluctuation in relevance (see figure 4). This is due to the default number of results displayed per page by each database: Academic Search Premier, n = 50; JSTOR and LexisNexis Academic, n = 25; ProQuest Central, PubMed, and Scopus, n = 20; Google Scholar and Web of Science, n = 10. As mentioned in the methodology, we only reviewed the first 25 results. Therefore, average scores did not change for Academic Search, JSTOR, or LexisNexis Academic. All but one of the other databases saw their average scores increase, most modestly, when only examining average scores from the first page. The largest increase was Google Scholar, with a 2.5 average that increased

---

**FIGURE 3**

Average of Interrater Pairs Rubric Scores by Database and Search Query Type

![Graph showing average rubric scores by database and search query type.](image)

**FIGURE 4**

What’s on the First Page (by Database)*

![Graph showing average rubric scores for the first page by database.](image)

*Numbers after database names indicate default number of results per page*
to 2.7 with only first-page results. PubMed was the only database whose average fell, 1.94 for all 25 results to 1.91 with only first-page results. A one-way between subjects ANOVA again confirms the changes were not significant.

When examining the mean and standard deviation of each rater for each database, standard deviation was lowest on either end of the spectrum (see table 5). Academic Search Premier had a lower standard deviation and some of the highest overall rubric scores, while ProQuest Central had a lower standard deviation but the lowest overall rubric scores. This also highlights the database that had the widest spread of rubric scores by rater. While almost all were consistent between raters, PubMed search results varied dramatically between Rater 1 and Rater 2, leading to disparate rubric scores between raters. Scopus searches, conversely, also had a significant lack of overlap in results between raters; however, this was not expressed as a difference in rubric score implying different articles of similar relevance.

While database averages tell us how well the database performed overall, the authors wanted to better understand the nuance of how each database searched (obviously without knowing the proprietary algorithms). To do this, we analyzed the overlap percentage by database, overlap percentage of filtered versus unfiltered results, and database precision.

Overlap percentage by database gives a sense of how different results are between searches as well as giving a sense for changes in database content (see figure 5). When each rater’s results were compared for their unfiltered Boolean versus unfiltered natural language searches, what was the overlap rate? A low overlap percentage means

<table>
<thead>
<tr>
<th>Database</th>
<th>Rater 1 Mean</th>
<th>Rater 1 Standard Deviation</th>
<th>Rater 2 Mean</th>
<th>Rater 2 Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Academic Search Premier</strong></td>
<td>2.48</td>
<td>0.510</td>
<td>2.48</td>
<td>0.510</td>
</tr>
<tr>
<td></td>
<td>2.48</td>
<td>0.510</td>
<td>2.44</td>
<td>0.507</td>
</tr>
<tr>
<td></td>
<td>2.64</td>
<td>0.569</td>
<td>2.64</td>
<td>0.569</td>
</tr>
<tr>
<td></td>
<td>2.65</td>
<td>0.489</td>
<td>2.65</td>
<td>0.489</td>
</tr>
<tr>
<td><strong>Google Scholar</strong></td>
<td>2.44</td>
<td>0.768</td>
<td>2.44</td>
<td>0.768</td>
</tr>
<tr>
<td></td>
<td>2.44</td>
<td>0.768</td>
<td>2.44</td>
<td>0.768</td>
</tr>
<tr>
<td></td>
<td>2.56</td>
<td>0.712</td>
<td>2.56</td>
<td>0.712</td>
</tr>
<tr>
<td></td>
<td>2.56</td>
<td>0.712</td>
<td>2.56</td>
<td>0.712</td>
</tr>
<tr>
<td><strong>JSTOR</strong></td>
<td>2.16</td>
<td>1.281</td>
<td>2.52</td>
<td>1.005</td>
</tr>
<tr>
<td></td>
<td>2.08</td>
<td>1.222</td>
<td>2.32</td>
<td>0.988</td>
</tr>
<tr>
<td></td>
<td>2.40</td>
<td>1.041</td>
<td>2.40</td>
<td>1.041</td>
</tr>
<tr>
<td></td>
<td>2.40</td>
<td>1.041</td>
<td>2.40</td>
<td>1.041</td>
</tr>
<tr>
<td><strong>LexisNexis Academic</strong></td>
<td>1.48</td>
<td>0.714</td>
<td>1.20</td>
<td>.707</td>
</tr>
<tr>
<td></td>
<td>1.32</td>
<td>0.945</td>
<td>1.48</td>
<td>0.714</td>
</tr>
<tr>
<td></td>
<td>1.48</td>
<td>0.714</td>
<td>1.48</td>
<td>0.714</td>
</tr>
<tr>
<td></td>
<td>1.48</td>
<td>0.714</td>
<td>1.48</td>
<td>0.714</td>
</tr>
<tr>
<td><strong>ProQuest Central</strong></td>
<td>1.92</td>
<td>0.862</td>
<td>1.08</td>
<td>0.812</td>
</tr>
<tr>
<td></td>
<td>1.48</td>
<td>0.770</td>
<td>0.52</td>
<td>0.586</td>
</tr>
<tr>
<td></td>
<td>1.92</td>
<td>0.862</td>
<td>1.48</td>
<td>0.770</td>
</tr>
<tr>
<td></td>
<td>0.52</td>
<td>0.586</td>
<td>0.52</td>
<td>0.586</td>
</tr>
<tr>
<td><strong>PubMed</strong></td>
<td>2.52</td>
<td>0.714</td>
<td>2.48</td>
<td>0.963</td>
</tr>
<tr>
<td></td>
<td>1.28</td>
<td>0.737</td>
<td>1.48</td>
<td>0.963</td>
</tr>
<tr>
<td></td>
<td>2.20</td>
<td>1.000</td>
<td>2.20</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>2.28</td>
<td>0.980</td>
<td>2.28</td>
<td>0.980</td>
</tr>
<tr>
<td><strong>Scopus</strong></td>
<td>2.04</td>
<td>1.020</td>
<td>2.04</td>
<td>0.935</td>
</tr>
<tr>
<td></td>
<td>2.32</td>
<td>0.900</td>
<td>2.00</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>2.32</td>
<td>0.918</td>
<td>2.00</td>
<td>1.000</td>
</tr>
<tr>
<td><strong>Web of Science</strong></td>
<td>2.00</td>
<td>1.225</td>
<td>2.08</td>
<td>1.115</td>
</tr>
<tr>
<td></td>
<td>2.28</td>
<td>0.980</td>
<td>2.28</td>
<td>0.980</td>
</tr>
<tr>
<td></td>
<td>2.28</td>
<td>0.980</td>
<td>2.28</td>
<td>0.980</td>
</tr>
</tbody>
</table>
results are closer to being unique; a user would find more relevant results by doing both searches. The 12 percent overlap rate for Academic Search Premier is quite low; a user might benefit from doing both a Boolean and natural language search. However, the overlap rate was similar for both raters, indicating database content was stable and raters found similar results. This is also true for Google Scholar and Web of Science. The wide difference in overlap rates for rater 1 versus rater 2 in both PubMed and Scopus is interesting and indicates frequently updated content. The timing of a search explains some of the discrepancies in overlap between raters and within rater searches. In general, overlap analysis indicates that natural language and Boolean searches are finding different (unique) results, suggesting researchers looking for all articles on a topic should do both searches to increase recall.

Something interesting happens when we examine only results that scored a 3 = very relevant on the rubric. Again, comparing the overlap of the unfiltered searches between Boolean and natural language syntax and remembering that a low overlap percentage means results are unique, most databases are above 50 percent, with Academic Search Premier more than 90 percent (see figure 6). This means that, for very relevant results, the search type does not matter (unfiltered Boolean or unfiltered natural language), the same very relevant results were found. For first-year students who struggle with constructing Boolean searches, they will retrieve the same highly relevant results with a natural language search. In other words, it would not be necessary to use Boolean to find relevant results in Academic Search Premier.

So, if highly relevant results have a lot of overlap, thus negating the necessity of introductory students using Boolean, is there a reason to use filters in a search? When examining overlap percentages, do filters really make a difference in the uniqueness of the results? Here, a high overlap percentage indicates high duplication between searches and would mean the filter is unnecessary (see figure 7). Note that Google Scholar overlap is not available because filtered results were identical to unfiltered results. PubMed, Scopus, and Web of Science have very high overlap percentages, indicating filtering has little to no effect on the diversity of the results. A low overlap percentage means the filter is causing more unique and relevant results to be included in the search results. In practice, the utility of filters is most apparent in ProQuest Central and Academic Search Premier, two of the most commonly taught databases for first-year
students. Filters also had some effect in JSTOR and LexisNexis Academic. Interestingly, the use of filters in Academic Search Premier had more effect with Boolean searches than natural language searches. Because of the disproportionate percentages between search query style, these findings suggest a divergent approach to instruction with that database in particular.

We next looked at the unfiltered searches to compare their precision. Precision is the number of relevant or very relevant articles (scoring either a 2 or 3 on the rubric) divided by the total number of citations retrieved (n = 25). Overall, results were similar between unfiltered Boolean and unfiltered natural language with the exception of ProQuest Central, which had an almost 20-point advantage of Boolean over natural

FIGURE 6
Overlap Percentage by Database—Results Scoring 3 on the Rubric (Unfiltered Boolean vs. Unfiltered Natural Language)

FIGURE 7
Overlap Percentage by Database (Unfiltered vs. Filtered Boolean and Unfiltered vs. Filtered Natural Language)
language (see figure 8). In half of the databases, natural language searches were slightly more precise than Boolean. In the remaining half, Boolean searches were more precise than natural language searches in three, and there was a tie in the case of PubMed. This indicates that either style of searching retrieves relevant results.

**Discussion**

Do Boolean searches retrieve more relevant results than natural language searches? For the majority of databases included in this study, both Boolean and natural language searching delivered results of highly comparable relevance. For both types of searches, the average relevance as well as the precision ratio of search results within each database (with the exception of ProQuest Central) was strikingly similar. “First-year” databases Academic Search Premier and Google Scholar performed exceptionally well. The variability in both the average relevance and the precision ratio of search results in ProQuest Central suggests that use of Boolean has some advantage over natural language searching for this particular database. For all other databases, there was no benefit to using Boolean over natural language searching in retrieving relevant results. This was the case when considering both the first twenty-five search results and the first page of results.

There is variation within databases. In general, Boolean and natural language searches yielded different (unique) results within a given database. Among the top 25 search results within each database, the overlap of Boolean and natural language search results was quite small regardless of whether or not the results were relevant. A student performing a Boolean search would find different results from a student using natural language even though average relevance would be similar. The overlap percentage of highly relevant results (those scoring a 3 out of 3 on the rubric), varied widely by database. This suggests that, for Google Scholar, Scopus, and Web of Science, it would benefit the searcher to do both a Boolean and natural language search if he or she wanted to get all of the most relevant results. However, for other databases, like Academic Search Premier, ProQuest Central, and PubMed, a searcher would be able to locate the same highly relevant results by using either of the two search strategies.

How accurate are database relevance ranking algorithms (on and beyond the first page) and should librarians be concerned if students do not go beyond the first page? First-year and introductory research assignments might typically ask students to locate only a
handful of relevant articles on a given topic. Based on this study, one page of results, and certainly the first 25, may be enough to satisfy the needs of a first-year research assignment. We found little difference in relevance further down the page, with the top third of results slightly outperforming the bottom third. This is helpful for teaching librarians to know, especially, as mentioned, studies have found the majority of students do not look past the first page of search results nor do they critically evaluate results further down that first page. This is not surprising given that convenience is a major factor in the way students search for information. The relevance of the first third of the first page of results provides evidence that librarians may not need to be too concerned about first-year students stopping after examining the first-page (or first half of the first page) of results.

Is there any advantage to using filters such as the scholarly article limiter? The high overlap percentages in PubMed, Scopus, and Web of Science indicate filtering has little effect on the diversity of the results. JSTOR and LexisNexis Academic were in the middle, indicating filters may have some effect. For teaching librarians, the low overlap percentage in first-year databases ProQuest Central and Academic Search Premier, Boolean search, indicate filters had a greater effect than in other databases tested. However, as stated previously, the high overlap for highly relevant results (rubric level 3) in Academic Search Premier indicate that, regardless of search type, first-year students would not necessarily notice that nuance.

Long live the Boolean? In what circumstances might Boolean be beneficial for undergraduate students? While outside the scope of this paper, Boolean may be more important for upper-level students or students with more complex research needs. For example, a student conducting an extensive literature review aiming for high recall (such as locating all possible relevant articles on a topic) or a student in a specialized discipline such as business where there may be many interrelated factors to consider (for example: industry, stakeholders, NAICS codes).

The Framework for Information Literacy for Higher Education does not explicitly address teaching Boolean operators, databases, or other technologies. In the past, the Information Literacy Competency Standards for Higher Education included Boolean operators as a performance indicator of a student who constructs and implements effectively designed search strategies. The Framework only implies the use of Boolean in the context of a learner using different types of searching language, a knowledge practice of the Frame “Searching as Strategical Exploration.” Regardless of whether the omission of the term Boolean operator in the Framework is significant, its absence allows librarians to decide their own approaches. Just as the Framework asks learners to be more critical in their thinking, it also asks librarians to be flexible and creative. It may also be worth noting that database technology and search engine optimization through algorithms and other means has changed significantly since ACRL published the Standards in 2000. As technology continues to change, librarians must continually evaluate how to use it to be most effective in their instruction.

One limitation of this study is that librarians did it without the direct involvement of student searchers. Although the search questions came from actual student research questions received by the authors, there is the possibility of benevolent bias. A student-centered study would add to this topic and is something the authors are exploring. Another limitation is that the present study only explored the Boolean operator AND. Additional research on the performance of the operators NOT and OR versus natural language is warranted. An additional area of study is to examine in greater depth exactly when, and under what circumstances, Boolean should be taught especially within the context of the Framework for Information Literacy.
Conclusion
While there are differences between and within databases, overall, a first-year, introductory student using virtually any database may not need instruction on Boolean search logic, as there is no clear advantage to Boolean in terms of retrieving relevant results on a topic. This is especially true for “first-year” databases such as Academic Search Premier and Google Scholar.

As discussed, it can take a considerable amount of class time to teach Boolean logic, and it is difficult for first-year, introductory students to understand and implement correctly. This study found no clear advantage in relevance of results between natural language and Boolean searching, suggesting that, for introductory courses, librarians can spend less time covering the mechanical “how to” aspect of searching and more time on other, more substantial information literacy concepts such as topic and question development (including search terms and terminology) and source evaluation. This approach supports the literature and the Framework, where teaching librarians express a desire to have more class time to get at those meatier concepts and research on the Framework indicates lower-level students can grasp these concepts. Topic development and source evaluation in particular stand out, as 84 percent of students in a 2010 Project Information Literacy study stated that getting started was the hardest part of the research process, and a 2016 Stanford study found students had difficulty evaluating information they found on the web. Searching is important, and teaching students to search is important, but these results demonstrate that teaching librarians can transition to focusing on more complex issues related to searching. For example, the thought process behind choosing search terms rather than the intricacies of how to link them together in a database.

Notes


17. Losee and Paris, “Measuring Search-Engine Quality and Query Difficulty,” as started in legal databases in the ‘90s. See, for example, Steve Arnold and Lawrence Rosen, “Bye, Boolean:


25. Dempsey and Valenti, “Student Use of Keywords and Limiters.”


36. While Google Scholar is a search engine, for the purposes of this study we refer to it as a database.


40. Asher, Duke, and Wilson, “Paths of Discovery”; Georgas, “Google vs. the Library (Part II).”

41. Georgas, “Google v the Library (Part II),” 524.

42. There was a significant effect of unfiltered versus filtered (Boolean and natural language) at the $P < .05$ level for the conditions $F(2,22) = 5.613, P = .011$. A follow-up independent-samples t-test was conducted to compare ProQuest filtered versus unfiltered Boolean and filtered versus unfiltered natural language. Both were significant at $P < .05$.

43. Note: due to differences in how results were recorded by interrater pairs, the overlap analysis was not able to be calculated for one reviewer for JSTOR, LexisNexis Academic, and ProQuest Central. Because content in some databases changes daily, it was impossible to recreate the searches. However, because the overlap analysis examined overlap between a single reviewer’s results, and not results between reviewers, it still provides evidence of the regularly changing database results.

44. Ibid.

45. Ibid.

46. Georgas, “Google v the Library (Part II)”; Asher, Duke, and Wilson, “Paths of Discovery.”

47. Georgas, “Google v the Library (Part II).”

48. Connaway, Dickey, and Radford, “If It Is Too Inconvenient.”


50. For example, Cmor and Li, “Beyond Boolean.”

51. For example, Dempsey and Valenti, “Student Use of Keywords and Limiters.”


53. Scott, “Part 1. If We Frame It”; Scott, “Part 2. If We Frame It.”
