

Oregon State University Libraries' Web of Science vs. Scopus Qualitative Comparison

Hannah Gascho Rempel, Associate Professor and Science Librarian

Report compiled for Oregon State University Libraries

June 2017

Overview

In Spring 2017, I was tasked with analyzing and providing feedback on the Web of Science and Scopus databases from an advanced researcher point of view. At Oregon State University (OSU), we currently have a subscription to Web of Science, and we wanted to see if switching to Scopus might be feasible from a number of different angles. The angle that I explored was whether advanced researchers might experience a qualitative difference when using one of these databases as compared to the other. To get at that question, I created two core searches for each of five disciplines that are most representative of the scholarship of OSU researchers. For each core search I performed a basic, relevance, advanced, and author search in both Web of Science and Scopus. A total of forty searches were performed. For each of those searches, I tracked a variety of factors that are important to searchers including overall number of results, date range of results, number of results with full-text available at OSU, perceived relevance of results, source types represented, and the number of overlapping results between the two databases. In addition, I observed interface features and interface similarities and differences between the two databases. Based on these searches, I determined that both Web of Science and Scopus provide highly relevant results on a wide disciplinary range of topics, but the results from the two databases are different. Both databases provide faceted searching, advanced analysis of results, and researcher metrics. Overall, either database would be a beneficial research tool for OSU researchers regardless of discipline. This report details the search strategies used, along with the results of the searches by discipline. A link to the raw data is included for those wishing to explore the data further.

Methods

Search Strategies Used

During April and May 2017, four main search strategies were used in both the Web of Science and Scopus databases. The following four search strategies were used:

1. Basic search – all default settings used, search box settings were left at either keyword topic (Web of Science) or article, abstract, title search (Scopus); search results were sorted by reverse date display, which was the default for both databases.
2. Relevance sort - all default settings used as described for the Basic search; search results sorted by relevance in both databases.

3. Advanced search – Boolean operators, additional synonyms, and quotes were used as appropriate for the search string to either broaden or narrow the search as necessary; search results were sorted by relevance. The same advanced search was used in both databases.
4. Author – based on the basic search, the first author of the top-cited paper was selected. Then a basic author search was conducted using the drop-down menu to the right of the search bar. While some advanced researchers may use a citation search, for the purposes of this exercise, simply recognizing that different fields could be selected to coincide with a particular query was deemed to be a more realistic representation of most advanced searchers' behavior based on my observations as a liaison and teaching librarian.

Searches Used by Discipline

Two core searches were conducted using each of the four search strategies described in the Search Strategies section in each of five disciplines. A total of 40 searches were conducted. The five disciplines chosen represent areas key areas of study at OSU. However, these disciplinary areas are quite common and have previously been used in other Web of Science vs. Scopus comparisons (Harzing and Alakangas 2016). Other researchers have either explored differences between the two databases in a single field (Abrizah et al. 2013; Kulkarni et al. 2009; López-Illescas, de Moya-Anegón, and Moed 2008; Meho and Rogers 2008). The search strings were selected based on a combination of my experience fielding questions from faculty and graduate students as a liaison librarian. For areas in which I had less liaison experience, I explored OSU faculty websites to find current research areas of interest. While these search strings are in no way a comprehensive representation of all the research that happens at OSU, let alone at other research institutions, the 40 searches did provide a qualitative snapshot of some of the similarities and differences between the two databases. The specific details of the search strings is included in the following tables. Note that the relevance sort search strategy is not included because the keywords used for that search were identical to the basic search, but instead of being sorted by reverse-date, those searches were sorted by relevance order.

Engineering Searches*

Search Type	Terms Used
Basic	<ol style="list-style-type: none"> 1. Bio-inspired sensing robots 2. Polymeric composites adhesion stress
Advanced	<ol style="list-style-type: none"> 1. ((bio-inspired OR soft robotics) AND (chemical OR neuro*) AND sensing) 2. natural polymeric composites adhesion (stress OR Strain)
Author	<ol style="list-style-type: none"> 1. Kernbach S* 2. Sreekala MS

*Information for the relevance sort searches is not included in any of these tables because the keywords entered were the same as the Basic search.

Health Science Searches

Search Type	Terms Used
Basic	<ol style="list-style-type: none"> 1. Environmental stimuli migraine

	2. Insulin resistance childhood obesity
Advanced	1. enviro* AND stimul* AND (migraine OR headache) 2. "childhood obesity" AND asthma AND ("insulin resistance" OR diabetes)
Author	1. Maggi CA 2. Weiss, Ram

Humanities Searches

Search Type	Terms Used
Basic	1. Personal epistemology reading 2. Feminist pragmatism
Advanced	1. "personal epistemology" reading 2. ((pragmatism OR postmodern*) AND (femin* AND discrim*))
Author	1. Schommer M* 2. Schilling C*

Science Searches

Search Type	Terms Used
Basic	1. Restored wetland salmon 2. Grain cellulose pectin
Advanced	1. anadromous habitat restor* 2. taurochenodeoxycholic OR taurochenodeoxycholate OR taurocholic OR taurocholate
Author	1. Simenstad CA 2. VansOest PJ

Social Science Searches

Search Type	Terms Used
Basic	1. Adolescent inmates recidivism 2. Food insecurity ethnography
Advanced	1. (adolescents OR juvenile) AND (inmates OR prisoners) AND (recidivism OR rehabilitation) 2. ("food insecurity" OR "food stress" OR "food poverty") AND (ethnograph* OR narrative OR phenomenological OR "case study")
Author	1. Freudenberg N* 2. Mintz SW

Criteria Tracked in Research Results for Both Databases

Based on the results returned for each of the search strings listed above, I used the following criteria in a spreadsheet to help guide my observations about the differences between the two databases. The criteria represent a range of qualitative and quantitative information to help provide insights into the underlying differences in breadth and precision of the two databases. Some of these metrics have been used in previous published comparisons of the two databases. For example, Archambault and collaborators looked at resource type differences (2009). More commonly, researchers have explored issues like citation counts, citation analyses, the number of articles in the two databases, and h-index

((Archambault et al. 2009; Harzing and Alakangas 2016; Vieira and Gomes 2009). Because I was primarily interested in how an individual advanced researcher might experience the two different databases, instead of focusing on larger publication metrics, I looked at issues that might impact a researcher's perception of the quality of the database. The criteria observed for each of the search strategies is outlined below.

Basic, Advanced, and Author Searches

- Overall number of results
- Date range of results
- How many of the top ten results are available at OSU?
- What is the perceived relevance of the top ten results? For the basic and advanced searches, relevance was determined by reading the title, abstract, or assigned subject headings to see if the results matched any of those fields. For the author searches, relevance was based on whether or not the author of the sources was the same author who wrote the initial top cited article. The relevance ranking was as follows: Very high (8-10 relevant results); High (6-7 relevant results); Moderate (4-5 relevant results); Low (2-3 relevant results); Very Low (0-1 relevant results).
- How many overlapping results are there between the two databases in the top ten results?
- How many of the top five journals in the list of results overlap between the two databases?
- Is the top cited article the same in both databases?
- If the top cited article is the same in both databases, how many times has that article been cited in each database?
- How many articles, book chapters, books, reviews, and conference proceedings are in the list of results?

Relevance Sort Search

Because the Relevance Sort search strategy was the same as the Basic search for several of the criteria, only the following criteria were tracked for the Relevance Sort search strategy.

- How many of the top ten results are available at OSU?
- What is the perceived relevance of the top ten results? Relevance was determined by reading the title, abstract, or assigned subject headings to see if the results matched any of those fields. The relevance ranking was as follows: Very high (8-10 relevant results); High (6-7 relevant results); Moderate (4-5 relevant results); Low (2-3 relevant results); Very Low (0-1 relevant results).
- How many overlapping results are there between the two databases in the top ten results?

Findings

Overall number of results

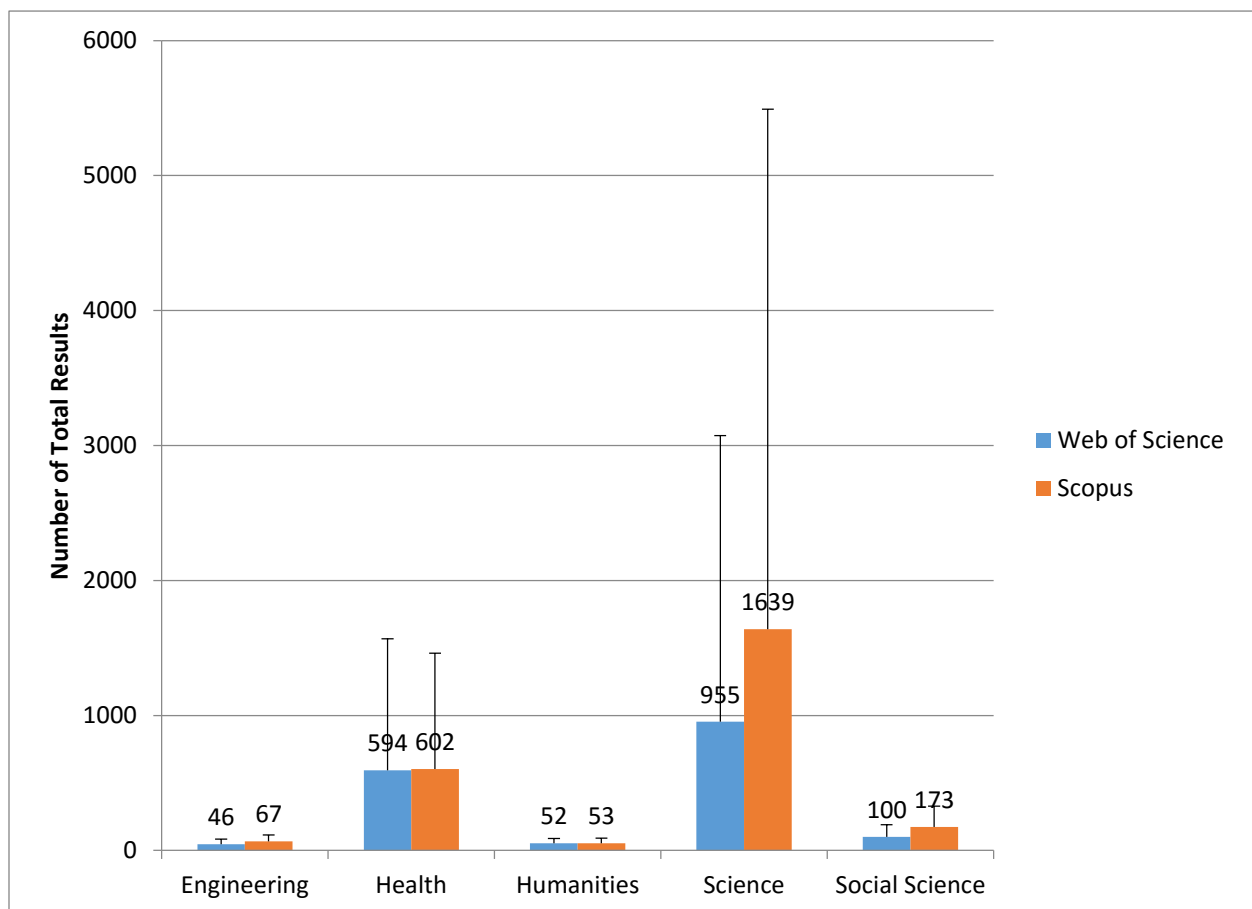
Because of the variation in searches between the basic, advanced, and author searches, statistically there was no significant difference in the number of overall results by database or discipline (see Table 1

and Figure 1). There were slight trends for more Engineering, Health, Science, and Social Science results in Scopus. There was virtually no difference in the number of results between Scopus and Web of Science.

Table 1. Average total number of results per search for all search types by discipline

Average Total Number of Results by Discipline		
	Web of Science	Scopus
Engineering	46	67
Health	594	602
Humanities	52	53
Science	955	1639
Social Science	100	173

Figure 1. Average total results per search for all search types by discipline including standard deviation



Relevance of Search Results

The two databases were virtually identical in terms of their relevancy rankings (see Table 2). Both databases had highly relevant search results in the top ten search results. Results tended to be very

relevant regardless of whether the relevancy sort was applied. *Disclaimer* – I had the lowest expert ability to determine relevancy in the Engineering searches, but based on looking at titles, abstracts, and keywords, results in Engineering still appeared to be highly relevant.

Table 2. Count of relevancy rankings by database

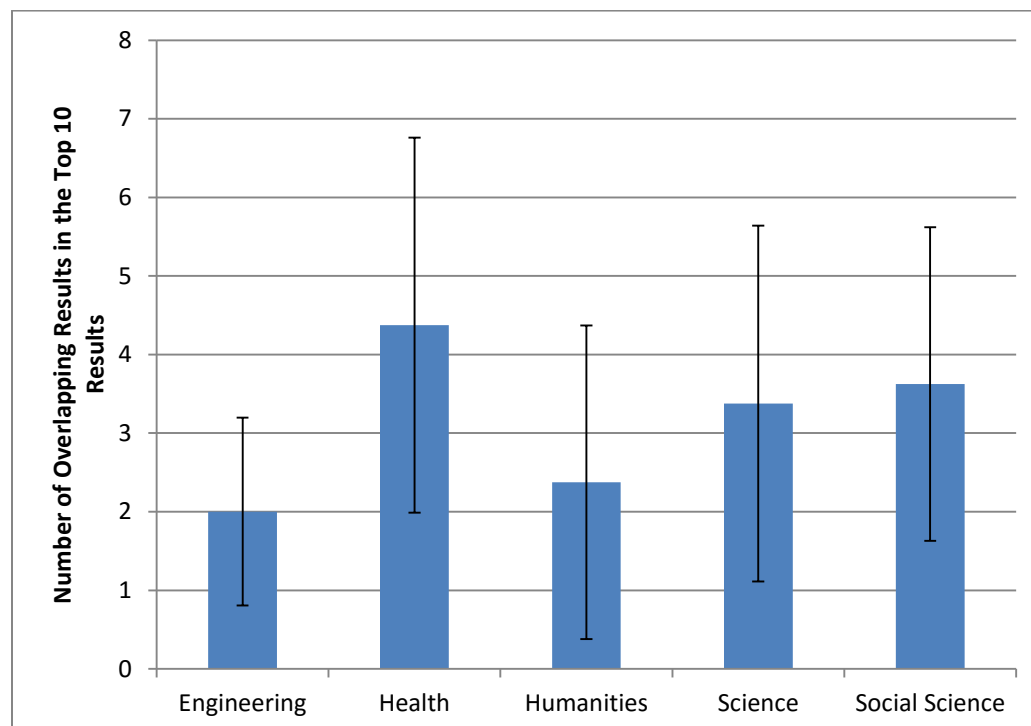
Count of Relevancy Rankings* by Database		
	Web of Science	Scopus
Very High	32	32
High	4	3
Medium	3	2
Low	1	2
Very Low	0	1

*Very high (8-10 relevant results); High (6-7 relevant results); Moderate (4-5 relevant results); Low (2-3 relevant results); Very Low (0-1 relevant results).

Number of overlapping results

Regardless of the discipline, there was fairly low overlap in the top ten results between the two databases (see Figure 2). There was no statistical difference between the disciplines, but Health Sciences tended to overlap the most at an average of 4.3 overlapping results per search; Engineering the least with 2 overlapping results per search. So while both databases had fairly similar and high relevancy rankings, the results themselves were different.

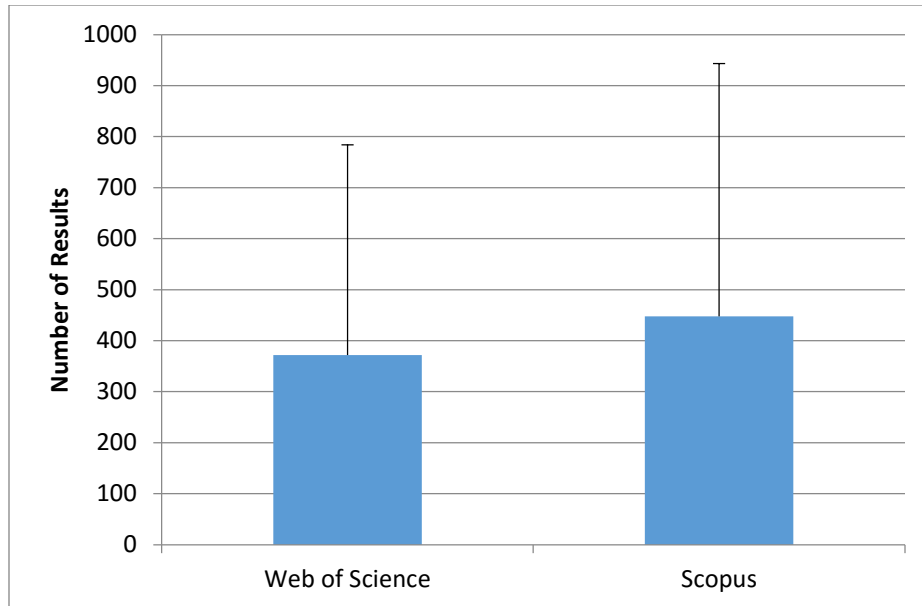
Figure 2. Average number of overlapping results by discipline



Times cited results

With the exception of one author publishing in the Health Sciences, Scopus consistently had more citations per highly cited work (see Figure 3). However, the difference is not enough to be significant.

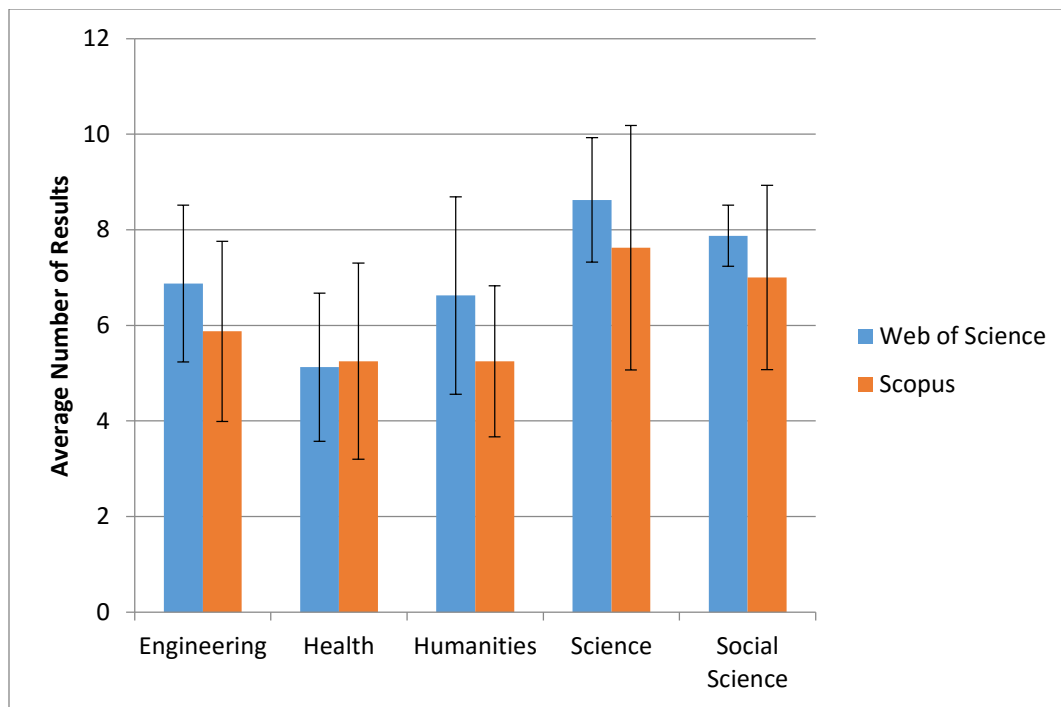
Figure 3. Average times cited results per database search when the top cited article was the same in both databases



Availability at OSU

In Scopus, there were slightly fewer search results from the list of top ten available at OSU for Engineering, Humanities, Science, and Social Science searches (see Figure 4). There were slightly more results available in Scopus at OSU for Health searches. None of these differences is significant.

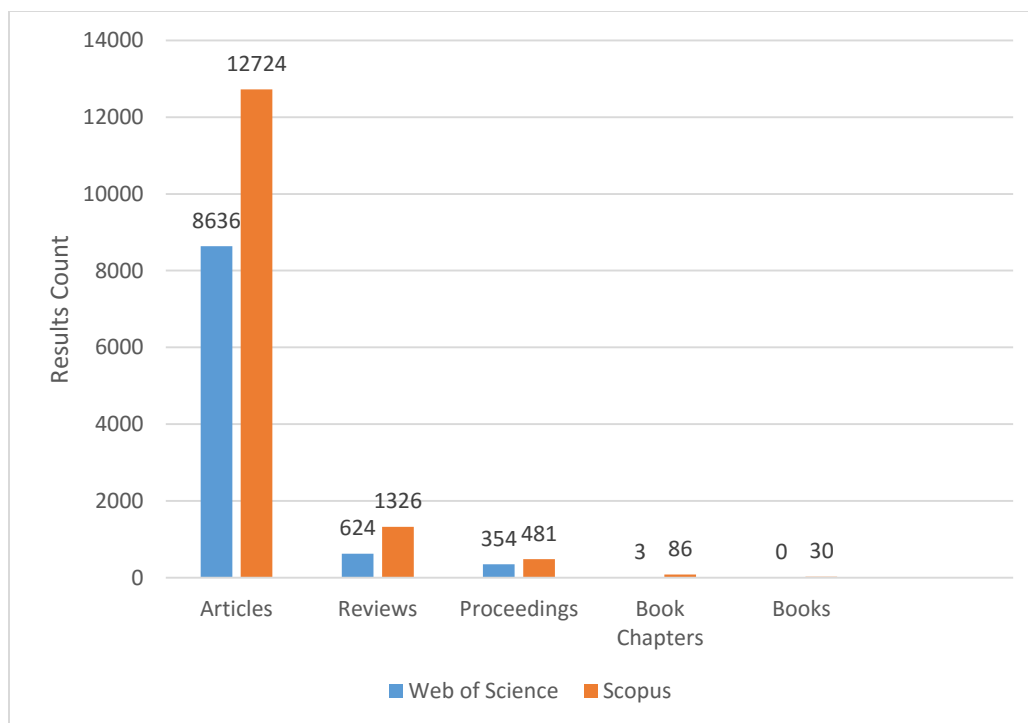
Figure 4. Average availability of top ten search results at OSU by discipline



Resource Type Differences

Overall, the higher number of total results in Scopus was primarily driven by a higher overall number of articles (see Figure 5). Beyond that difference, there also tended to be more review articles, somewhat more proceedings, and more book chapters available in Scopus. Notable differences by discipline (again, non-significant differences) were that in Scopus, for Engineering there were more proceedings in the results lists. In Scopus, for Health and in Social Science there were more review articles.

Figure 5. Total for each research type across all disciplines by database.



Interface Observations

Overall, many of the search interface features in the two databases were similar. Scopus doesn't have a funding facet, but once a list of results are exported, the funding information can be compiled in a CSV. Unlike Web of Science, Scopus does have a keyword facet, which could be helpful when demonstrating searching to undergraduates and new graduate students.

With Scopus a standalone bibliography can be created based on 10 common styles straight from the results list (it uses QuikBib for this feature).

Scopus has an Open Access notation clearly visible in the results list below each open access article.

Scopus tended to do a better job displaying correct author information when doing an author search and more clearly disambiguated authors with similar names.

Conclusions

Both Web of Science and Scopus provide highly relevant results on a wide disciplinary range of topics, but the results from the two databases are different. Both databases provide faceted searching, advanced analysis of results, and researcher metrics. Overall, either database would be a beneficial research tool for OSU researchers regardless of discipline.

Link to raw data file:

https://drive.google.com/file/d/0B_xcIEG6DYt7ZFROWEs4bWs5SDg/view?usp=sharing

References

- Abrizah, Abdullah, A. N. Zainab, K. Kiran, and R. G. Raj. 2013. "LIS Journals Scientific Impact and Subject Categorization: A Comparison between Web of Science and Scopus." *Scientometrics* 94 (2): 721–740.
- Archambault, Éric, David Campbell, Yves Gingras, and Vincent Larivière. 2009. "Comparing Bibliometric Statistics Obtained from the Web of Science and Scopus." *Journal of the American Society for Information Science and Technology* 60 (7): 1320–26. doi:10.1002/asi.21062.
- Harzing, Anne-Wil, and Satu Alakangas. 2016. "Google Scholar, Scopus and the Web of Science: A Longitudinal and Cross-Disciplinary Comparison." *Scientometrics* 106 (2): 787–804. doi:10.1007/s11192-015-1798-9.
- Kulkarni, Abhaya V., Brittany Aziz, Iffat Shams, and Jason W. Busse. 2009. "Comparisons of Citations in Web of Science, Scopus, and Google Scholar for Articles Published in General Medical Journals." *Jama* 302 (10): 1092–1096.
- López-Illescas, Carmen, Félix de Moya-Anegón, and Henk F. Moed. 2008. "Coverage and Citation Impact of Oncological Journals in the Web of Science and Scopus." *Journal of Informetrics* 2 (4): 304–316.
- Meho, Lokman I., and Yvonne Rogers. 2008. "Citation Counting, Citation Ranking, and h-Index of Human-Computer Interaction Researchers: A Comparison of Scopus and Web of Science." *Journal of the American Society for Information Science and Technology* 59 (11): 1711–1726.
- Vieira, Elizabeth, and José Gomes. 2009. "A Comparison of Scopus and Web of Science for a Typical University." *Scientometrics* 81 (2): 587–600.