Bibliometrics for Faculty Evaluation: A Statistical Comparison of h-indexes Generated Using Google Scholar and Web of Science Data

David Dunaway, Louisiana State University

Abstract

The growing need for quantification of research performance for promotion and tenure and grant funding decisions has lead many to rely on citation metrics. There are many metrics to choose from but one of the most common is the *h*-index. While the *h*-index has been criticized by many, the metric itself is not the only concern. The source of the citation information used to calculate the *h*-index is also important. In this case study the *h*-index was calculated using citation data from Clarivate's Web of Science (WoS) and Google Scholar (GS) for a selection of faculty working at a large public university. The *h*-indexes from the two sources were statistically compared using a student's t-test and Spearman correlation to determine if the two sources produced significantly different results. Google Scholar data produced *h*-indexes that were greater in magnitude (M=18.52, SD=13.641) than those produced by Web of Science data (M=13.13, SD=10.400) however the rank order of the *h*-indexes from the two sources showed a high degree of similarity.

Keywords: bibliometrics; *h*-index; scientometrics; Informetrics; research Evaluation

Researchers and their research are continuously evaluated for the purpose of funding by research grants and for promotion and tenure decisions. To make these decisions as objective as possible, especially in the STEM fields, many have sought a suitable system of numerical rating. The first use of citation counts in the evaluation of scientific work was published in a 1927 work by Gross and Gross where the authors were seeking an arbitrary standard for the selection of chemistry journals (Gross & Gross, 1927). Since that early beginning, citations have been analyzed for assessment of national science policies, departments and individual scientists (Bornmann & Daniel, 2008). Citations counts are considered valid measures of impact on the scientific community because it is assumed that higher quality work is cited more than lower quality work (van Raan, Visser, Van Leeuwen, & van Wijk, 2003). While counting total papers published or total citations can be used to evaluate research it is problematic because over a researcher's career this favors researchers that have been actively performing research and writing articles longer as articles gather more citations with time and would be more of a measure of quantity of work rather than quality (Ball, 2005) (Hargens, 2000).

There are many metrics that have been developed that employ some calculations involved with article citation information to gauge quality of research. One of the most popular is the *h*-index or Hirsch index (Hirsch, 2005). The *h*-index was developed by J. E. Hirsch, a physicist working at the University of California, San Diego. Hirsch proposed an index *h* defined as the number of papers with citation number $\geq h$. "In this system a scientist has index *h* if *h* of his or her N_p papers have at least *h* citations each and the other $(N_p - h)$ papers have $\leq h$ citations each" (Hirsch, 2005). This is a simple measure that combines both the number of articles or the quantity of publications, and the number of citations or the quality or visibility, of publications.

There have been many criticisms of the h-index. One consideration is that it is not sensitive to highly cited papers or papers with low or zero citations. The latter is an advantage of the h-index, but the former is problematic as highly cited papers should count as a measure of quality or visibility of a researcher's work. While this is supposed to distinguish a "one-hit wonder" from a steady performer (Cronin & Meho, 2006), it doesn't reflect the importance of early career researchers' work. For example, two researchers that both have written ten articles could have the same h-index even if one or more of one researcher's papers had a much higher number of citations while the other did not.

Table 1

Example h-index calculation

Researcher 1		Researcher 2	
Article	Citations	Article	Citations
1	35	1	10
2	25	2	9
3	20	3	8
4	7	4	6
5	**5**	5	**5**
6	4	6	4
7	2	7	3
8	2	8	2
9	1	9	1
10	1	10	0

Table 1

In the example above, both researchers would have an h-index of 5 even though researcher 1 had three well cited articles and researcher 2 did not. Even with this limitation, the h-index is used by many institutions and funding agencies to evaluate research and researchers. Hirsch went so far as to give specific values of the *h*-index for certain types of advancement in a physicist's career. An *h*-index of about 12 would be enough for achieving tenure at the associate professor level in major research universities and an *h*-index value of 18 for full professor. Membership in the National Academy of Sciences would require an *h*-index of 45 (Hirsch, 2005).

This might be a reasonable benchmark for physicists, but publication and citation rates vary across disciplines (Hurt, 1987) (Hargens, 2000). Kelly and Jennions indicate that cell biology publications accumulate citations at a greater rate than do ecological publications (Kelly & Jennions, 2006). Other disciplines vary as well. Vieira and Gomes rank Biology & Biochemistry highest in citations per article of the four disciplines studied followed by Chemistry and Physics while Mathematics was lowest (Vieira & Gomes, 2010). The data collected in this case study show that the College of Coast & Environment had an average *h*-index of 23 followed by Physics & Astronomy with 19 using WoS data whereas both departments have an average *h*-index of 27 using GS data. The data by department can be seen in Table 2.

Table 2

Average h-index by department

	Avg. h-index WoS	Avg. No. of Articles WoS	Avg. Citations per Article WoS	Avg. h-index GS	Avg. No. of Articles GS	Avg. Citations per Articles GS
Coast & Environment	23	113	28	27	117	33
Physics & Astronomy	19	62	34	27	143	29
Chemistry	17	37	38	19	59	30
Geology	16	35	34	25	134	15
RNR	16	24	35	25	148	18
Biological Sciences	14	43	29	18	96	19
Engineering	12	30	29	18	87	22
Math	10	38	8	16	64	13
Geography & Anthropology	8	19	10	14	79	11

Note that this is a very small sample and should by no means be interpreted as representative of the disciplinary population.

Table 2

Article type also affects citation rates. Review articles receive a disproportionate number of citations (MacRoberts & MacRoberts, 1996). This would indicate that if the *h*-index were used as a measure for tenure and promotion a scale would have to be developed for each individual discipline and the comparison of researchers across disciplines would not be equitable.

Other measures have been created to address these limitations such as the g-index proposed by L. Egghe (Egghe, 2006). The g-index is defined as the highest number g of papers that received g^2 or more citations. This means that g ≥h so that the g-index score will be higher for all articles and authors then the h-index (Egghe, 2006). What makes this index different is that the higher the number of citations in the upper articles, the higher the g-index. As was shown above, this is not the case with the *h*-index. Using the same example data above to calculate the g-index illustrates the difference.

Table 3

Example g-index calculation

Researcher 1				Researcher 2			
Article	Citations	∑ Citations	$Article^2$	Article	Citations	∑ Citations	$Article^2$
1	35	35	1	1	10	10	1
2	25	60	4	2	9	19	4
3	20	80	9	3	8	27	9
4	7	87	16	4	6	33	16
5	5	92	25	5	5	38	25
6	4	96	36	6	4	42	36
7	2	98	49	7	3	45	49
8	2	100	64	8	2	47	64
9	1	101	81	9	1	48	81
10	1	102	100	10	0	48	100

Table 3

In this example, Researcher 1 would have a g-index of 9 and Researcher 2 a g-index of 6. The higher number for Researcher 1 reflects the larger citation numbers for Researcher 1’s articles.

The source of citation information used to calculate the h-index is also very important. The most frequently used sources for citation information are Clarivate’s Web of Science, Elsevier’s Scopus and Google Scholar. At the author’s university, faculty in at least one of the departments in the college of science are required to include citation information in their annual review, including the h-index, from either Web of Science or

Google Scholar as the university libraries does not subscribe to Scopus. The purpose of this research is to compare citation information for researchers at the university from Web of Science and Google Scholar to determine if the two produce the same or similar values for the h-index.

Method

This is a case study of research article citations for research done at the author's university. This sample was chosen because at least one of the departments on campus requires faculty members to report their h-index for their annual evaluation but they can choose either Web of Science or Google Scholar as the source of their information. The author was interested in seeing if the h-index values generated by these two sources were equitable. In order to compile a list of h-index values from a list of publications as similar and complete as possible for each researcher, only researchers with both a Web of Science ResearcherID and Google Scholar profile were included in this research.

ResearcherID is an author identifier that allows researchers with access to WoS to create a publication list that is assigned a unique identifier that can be used to view all of a researcher's publications without having to search the database by author. Searching by author is especially problematic in WoS where the author' name is indexed as last name and first initial only. Author disambiguation is a problem in WoS as in other databases and the ResearcherID is designed to help alleviate this problem.

The Google Scholar profile allows researchers to create a similar list of publications that are curated by the author. Both the ResearcherID and Google Scholar profile are created and curated by the researcher therefore the lists of publications are authoritative. The researchers with ResearcherIDs were found by performing a search in

Clarivate's Web of Science (WoS) ResearcherID database using the university name in the 'Institution' field. This produced a list of all authors from the institution that published papers indexed by Web of Science that had created a ResearcherID.

This list was used to perform author searches in Google Scholar to determine which authors also had a verified Google Scholar profile. The ResearcherID was used to obtain the author's status, department, h-index, total number of articles, and total number of citations. The Google Scholar profiles were used to obtain the author's h-index, total number of articles, and total number of citations.

SPSS was used to perform a paired Student's t-test with the h-indexes generated by Web of Science and Google Scholar to determine if there was a statistical difference in the two lists. A 2-tailed Spearman's rank-order correlation was run to determine the relationship between the rank order of h-indexes generated by Web of Science and Google Scholar.

Results

The search of Web of Science for researchers at the author's institution with a ResearcherID returned 300 results. All 300 researcher names were then searched in Google Scholar yielding 54 researchers that had both a Web of Science ResearcherID and a Google Scholar profile. A paired Student's t-test was conducted to compare the h-index scores from both sources. There was a significant difference in the h-index scores from Web of Science ($M=13.13$, $SD=10.400$) and Google Scholar ($M=18.52$, $SD=13.641$); $t(53) = 6.293$, $p < 0.001$. These results suggest that h-index values generated by Google Scholar are higher than those generated by Web of Science for the same researchers.

A Spearman rank-order correlation was computed to assess the relationship between the *h*-indexes generated using WoS and GS. There was a strong, positive correlation between h-indexes generated by Web of Science and Google Scholar, which was statistically significant, r(54) = .997, n = 54, $p < .001$. This indicates that the rank order is very similar for WoS and Google Scholar ($r_s = 1.0$ indicates a perfect match).

Table 4

Spearman Correlation

	h-index WoS	h-index GS
h-index WoS		
Spearman Correlation	1	.997**
Sig. (2-tailed)		.000
N	54	54
h-index GS		
Spearman Correlation	.997**	1
Sig. (2-tailed)	.000	
N	54	54
** Correlation is significant at the p < 0.001 level (2-tailed)		

Table 4

Discussion

There are many criticisms for using the h-index for such important purposes as promotion and tenure decisions as well as grant funding. The results of this investigation indicate that although the h-index generated by Google Scholar data are larger than those generated with Web of Science, the rank order for the two lists were similar as indicated by the Spearman's rho Correlation value of $r_s = .997$. This shows that if all faculty are required to use the same source, either of the sources could be used for their evaluation information. If some researchers choose to use Web of Science while others use Google Scholar, the Google Scholar values will give the appearance of greater value to their research.

The disciplinary differences in publication and citation rates should be addressed to maximize the utility of the measure. Individual departments will need to develop a scale of what acceptable *h*-index values are for their specific disciplines. Those provided by Hirsch would only be useful for physicists and might need to be adjusted over time as citation rates have increased for many disciplines since the introduction of the *h*-index in 2005 (Smith, 2008). There is also research that shows that female scientists tend to publish significantly fewer publications and their publications are cited less than those of men (Aksnes, Rorstad, Piro, & Sivertsen, 2011). This could lead to a gender bias in funding and promotion and tenure.

Conclusion

Quantitative evaluation of research is important but problematic. Many metrics have been developed based on citation data that attempt to provide a measure of value or quality of research the most common being the h-index. The source of citation information can have a large impact on the value of citation metrics. The h-indexes generated with citation data from Google Scholar was consistently larger than those generated from Web of Science citation data. However, the rank order of the h-indexes generated from both sources was very similar. Either source could be used to rank research in a department as long as the same source is used for all researchers.

Limitations of the Study

The sample size is small and is take from faculty publications from one University. A larger sample size and a variety of Universities would make the results more representative of the population of all faculty and institutions. This is especially true of the discussion of individual departments as the sample sizes become very small.

References

Aksnes, D. W., Rorstad, K., Piro, F., & Sivertsen, G. (2011). Are female researchers less cited? A large-scale study of Norwegian scientists. *Journal of the American Society for Information Science & Technology*, *62*(4), 628–636. https://doi.org/10.1002/asi.21486

Ball, P. (2005). Index aims for fair ranking of scientists. *Nature*, *436*(7053), 900–900. https://doi.org/10.1038/436900a

Bornmann, L., & Daniel, H. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*(1), 45–80. https://doi.org/10.1108/00220410810844150

Cronin, B., & Meho, L. (2006). Using the h-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology*, *57*(9), 1275–1278. https://doi.org/10.1002/asi.20354

Egghe, L. (2006). Theory and practice of the g-index. *Scientometrics*, *69*(1), 131–152. https://doi.org/10.1007/s11192-006-0144-7

Gross, P. L. K., & Gross, E. M. (1927). College libraries and chemical education. *Science*, *66*(1713), 385–389. Retrieved from JSTOR.

Hargens, L. L. (2000). Using the literature: Reference networks, reference contexts, and the social structure of scholarship. *American Sociological Review*, *65*(6), 846–865. https://doi.org/10.2307/2657516

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569–16572.

Hurt, C. D. (1987). Conceptual citation differences in science, technology, and social sciences literature. *Information Processing & Management*, *23*(1), 1–6. https://doi.org/10.1016/0306-4573(87)90033-1

Kelly, C., & Jennions, M. (2006). The h index and career assessment by numbers. *Trends in Ecology and Evolution*, *21*(4), 167–170. https://doi.org/10.1016/j.tree.2006.01.005

MacRoberts, M. H., & MacRoberts, B. R. (1996). Problems of citation analysis. *Scientometrics*, *36*(3), 435–444. https://doi.org/10.1007/BF02129604

Smith, D. R. (2008). Citation analysis and impact factor trends of 5 core journals in occupational medicine, 1985-2006. *Archives of Environmental & Occupational Health*, *63*(3), 114–122. https://doi.org/10.3200/AEOH.63.3.114-122

van Raan, A. F. J., Visser, M. S., Van Leeuwen, T. N., & van Wijk, E. (2003). Bibliometric analysis of psychotherapy research: Performance assessment and position in the journal landscape. *Psychotherapy Research*, *13*(4), 511–528. https://doi.org/10.1093/ptr/kpg038

Vieira, E. S., & Gomes, J. A. N. F. (2010). Citations to scientific articles: Its distribution and dependence on the article features. *Journal of Informetrics*, *4*(1), 1–13. https://doi.org/10.1016/j.joi.2009.06.002