



# SAVVY SEARCHING

## The h-index for countries in Web of Science and Scopus

The h-index

831

Péter Jacsó

*University of Hawaii, Honolulu, Hawaii, USA*

### Abstract

**Purpose** – The purpose of this paper is to discuss the results of recent experiments in determining the h-index at the country level for the 10 Ibero-American countries of South America.

**Design/methodology/approach** – The three citation index components (Science Citation Index, Social Science Citation Index, and Arts & Humanities Citation Index) of the Web of Science system of Thomson-Reuters and the Scopus database of Elsevier are used to gauge the comparability of the h-indexes reported by the two systems.

**Findings** – The results show that in spite of the significant differences in the content of the two databases in terms of their source base and the extent of cited reference enhancement of records, the rank correlation of the ten countries based on the h-index values returned by Web of Science and Scopus is very high.

**Originality/value** – For this sample, in spite of differences between WoS and Scopus in the number of papers from each country, the rank position of the countries by the h-index is almost the same. There is only a single rank position difference – Scopus rank Argentina second and Chile third, while the order is the reverse in WoS. This reconfirms the robustness of the h-index at the country level.

**Keywords** Publishing, Databases, South America

**Paper type** Viewpoint

Originally, the concept of the h-index was developed to measure the scholarly publishing productivity and impact of individual researchers through a single indicator. It became widely popular for the simplicity of the concept, and its use was extended to research groups, institutions and journals. This paper discusses the results of my recent experiments in determining the h-index at the country level for the ten Ibero-American countries of South America. The three citation index components (Science Citation Index, Social Science Citation Index, and Arts & Humanities Citation Index) of the Web of Science system of Thomson-Reuters and the Scopus database of Elsevier were used to gauge the comparability of the h-indexes reported by the two systems. The experiments covered the 1981 to 2007 publication years as the time period. The citation window was extended to mid-January 2009. The results show that in spite of the significant differences in the content of the two databases in terms of their source base and the extent of cited reference enhancement of records, the rank correlation of the ten countries based on the h-index values returned by Web of Science and Scopus was very high.



## Introduction

The h-index was introduced by Jorge E. Hirsch (2005) to quantify the scientific publication output and the impact of the work of researchers. It is a composite measure based on the combination of the number of papers published and the number of citations these papers have received according to records created for and reported by Web of Science (WoS) and Scopus.

The h-index was originally meant to measure the scientific performance of researchers through the prism of publications. It received much attention and wide acceptance from the most respected scientometricians, along with several suggestions and recommendations for derivative indexes (Bornmann and Daniel, 2007; Egghe, 2006, Glänzel, 2006; Rousseau and Ye, 2008; Schreiber, 2007; Schubert and Glänzel, 2007. Bornmann and Daniel (2009) has undertaken the most current review of research activities related to the h-index). Quite soon the index was extended to measure the scholarly productivity and impact of journals and of universities and other research institutions and groups (Braun *et al.*, 2005; Levitt and Thelwall, 2009; Meneghini and Packer, 2006; Prathap, 2006; Van Rann, 2006). Many scholars have argued for and used the h-index for ranking researchers and research groups in a specific field or country (Cronin and Meho, 2006; Meho and Rogers, 2008; Meneghini and Packer, 2006; Molinari and Molinari, 2008; Oppenheim, 2007). Some researchers have analysed the content and software features of the systems and services most widely used for determining the h-index (Bar-Ilan, 2008; Jacso 2008a, b).

It is natural to also contemplate the use of the h-index for assessing the scholarly publishing and thus scientific research at the country level. Other bibliometric indicators have been used for this purpose for a long time (King, 2004; Moravcsik, 1985).

## Objectives and methodology

The purpose of this research was to test the feasibility of using WoS and Scopus to determine the h-index at the country level in order to create rank lists of countries from the perspective of research productivity and impact, and to explore topic-specific issues, such as the extent of availability of authors' country affiliation in the records, the consistency and accuracy of the format of the assigned country codes/names, the handling of the exceptionally large sets that can be produced when searching by country as a sole search criterion (that is, without limiting the search by a disciplinary category, topic, author names or combination of other search terms) and, most importantly, the plausibility and comparability of h-indexes generated from WoS and Scopus. (Google Scholar does not offer country-specific index and search options, and even if it did, they would be of no more use than the other field-specific indexes purportedly available but practically dysfunctional, such as the broad subject category codes and publication years.) This paper presents the findings for the ten Ibero-American countries; the other findings will be discussed in a separate paper.

The countries of South America were chosen primarily because there are distinct groups of countries on the continent with high, medium and low scientific activity and publishing productivity (in line with the differences in their overall economic position, tertiary education and research opportunities and resources). Brazil, Argentina and Chile are at the top end of the scale, while Ecuador, Bolivia and Paraguay are at the

bottom end of the scale, with the other four countries in between (see Tables I and II). The geographic and economic strata offered a natural clustering for testing the flexibility and feasibility of using the h-index for countries of very different scientific profile, and for exploring the pros and cons of WoS and Scopus from the more widely applicable methodological perspective.

The time period used in the investigation was limited to the 1981 to 2007 publication year range. The citation window was longer to accommodate citations received to the very end of 2008. This time frame was a compromise between too short and too long a window, and provided a level playing field for Scopus and WoS to put their best foot forward – Scopus with its broader source coverage since about 2000, and WoS having cited reference enhanced records for its entire database, irrespective of the time span, not just from 1996 as with Scopus.

It was an additional advantage that the 1981 to 2007 segments of the two databases were of similar size – 28,690,676 master records in WoS and 27,857,435 master records in Scopus. Searching was especially easy, as none of the countries had merged, separated or changed their status or their names between 1981 and 2007, and none of them has a name that appears with different abbreviations, formats or spelling within and among the databases. These are important practical issues in searching by

WoS Country	H-I	Only citable documents			Any documents		
		Rank	Papers	Rank	H-I	Rank	Papers
Argentina	175	3	76,914	2	175	3	100,141
Bolivia	47	9	1,556	9	48	9	1,929
Brazil	214	1	186,223	1	210	1	243,240
Chile	178	2	40,210	3	180	2	52,733
Colombia	101	5	10,022	5	103	5	13,145
Ecuador	60	8	2,100	8	61	8	2,547
Paraguay	30	10	435	10	31	10	606
Peru	81	7	4,499	7	82	7	6,336
Uruguay	84	6	4,983	6	85	6	6,300
Venezuela	119	4	18,058	4	120	4	23,267

**Table I.**

Results from WoS,  
limiting the search to  
“citable” document types  
and searching for any  
type of document

Scopus Country	H-I	Only citable documents			Any documents		
		Rank	Papers	Rank	H-I	Rank	Papers
Argentina	157	2	81,690	2	158	2	85,722
Bolivia	41	9	1,490	9	41	9	1,545
Brazil	216	1	229,186	1	218	1	238,630
Chile	142	3	38,171	3	142	3	39,768
Colombia	86	5	12,882	5	86	5	13,388
Ecuador	55	8	2,294	8	56	8	2,498
Paraguay	29	10	468	10	30	10	510
Peru	68	7	4,308	7	69	7	4,556
Uruguay	79	6	5,467	6	79	6	5,766
Venezuela	103	4	19,888	4	103	4	20,745

**Table II.**

Results from Scopus,  
limiting the search to  
“citable” document types  
and searching for any  
type of document

country. Being unaware of them can significantly distort the results, but these are administrative rather than substantial aspects for this research.

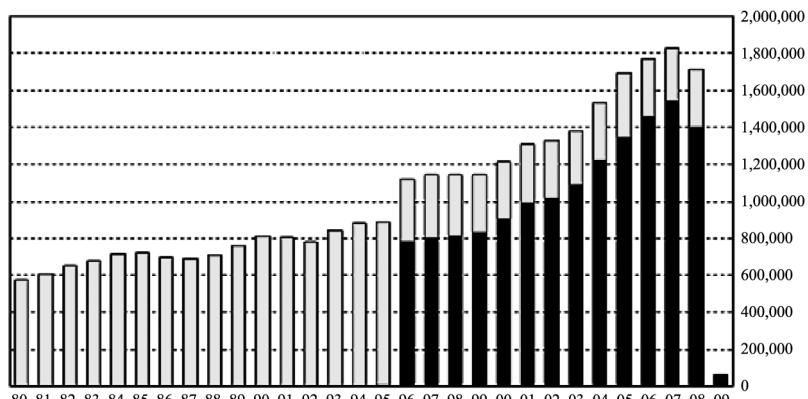
There are three essential content differences between WoS and Scopus from the perspective of searching by country as the primary criterion. First, in Scopus 12.5 million records (34 per cent) had no country name in the AFFILCOUNTRY field and there were more than 3.5 million such records in the chosen subset. In WoS, the omission rate ranged from 12 per cent to 15 per cent for records between 1981 and 2007, while in the chosen subset for this research, the author estimates that slightly more than 3 million records had no country affiliation in WoS. As WoS currently does not offer a search option to gauge this value, the estimate is based on searches made on the Dialog implementation of the Thomson-Reuters citation databases, using the Geographic Location (GL) field.

Another important difference that affects every type of search is that Scopus has added cited references to its records only since 1996 (and to about 7,000 records for the period before). WoS has enhanced all its records, with cited references present in the source documents from the beginning. Figure 1 illustrates the presence of cited reference enhanced records in Scopus.

A third trait can be important for those who wish to limit the country search by document type to journal articles, review papers and notes (and in Scopus also to conference papers). While WoS has a document type field in every record, in Scopus more than 1.5 million records in the 1981 to 2007 subset had no document type information. For the entire database range of 1823-2008, Scopus had 3.3 million records with no document type assigned.

### Findings

The results of the series of searches produced plausible h-index values for all the countries. Table I shows the results from WoS when the search was limited to documents considered to be the primary target of citations, such as reviews, journal articles and research notes, and excluding document types that are often referred to as "non-citable documents" (letters to the editor, book reviews, corrections, etc.). There are practically no differences between the results of the two types of searches, unfiltered and limited to citable items, respectively – except for a minor one in the case of Brazil.



**Figure 1.**  
Cited reference enhanced records in Scopus are shown in dark

The same symptom was evident for the two types of searches in Scopus, where even for Brazil the difference was merely 2 points, and for the rest, only 1 point or none at all. The same intra-database pattern was found by this author when calculating the h-index for Australia and New Zealand.

The searches were first performed without document type filtering, in order to avoid penalising Scopus for the missing document types in more than 1.5 million records. After the positive results of the first round of searching, the test was repeated with the “citable document” filter (see Tables I and II). Later it was found that Scopus adds the “undefined” category in an OR relationship to the document type term supplied by the user. This eliminates the handicap caused by the omission of document types. Time did not allow a rerun of the searches by removing the extra term, “undefined”, and re-tabulating the results. This will be done in the next round of testing country-level scholarly publishing and productivity indicators. WoS has document type assigned for all records, and its own searches indicate that this distinction did not matter for the record sample for these countries. Lack of document type assignment makes significant differences for the traditional citation rate indicator. If records with “non-citable” and without assigned document types are removed, then the denominator is reduced and the citation rate increases. This may lead to unfair practices by the publishers, who want to increase their journal’s citation rate.

The lack of cited references in the records of pre-1996 publications did not have the assumed negative effect on the h-index – presumably they were offset by the intensive enhancement of the source coverage in the past few years.

## Conclusions

For this sample, in spite of differences between WoS and Scopus in the number of papers from each country, the rank position of the countries by the h-index was almost the same. There was only a single rank position difference – Scopus ranked Argentina second and Chile third, while the order was the reverse in WoS. This reconfirms the robustness (Rousseau, 2007; Vanclay, 2007) of the h-index at the country level. For this sample, the h-index comes as close to consensus as is reasonable when two different systems are used with considerable differences from the traditional database content evaluation and information retrieval perspective. The results are reassuring about the viability of using the h-index for purposes of measuring and ranking the scientific performance and impact of countries.

## References

- Bar-Ilan, J. (2008), “Which h-index? A comparison of WoS, Scopus and Google Scholar”, *Scientometrics*, Vol. 74 No. 2, pp. 257-71.
- Bornmann, L. and Daniel, H.-D. (2007), “What do we know about the h index?”, *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 1, pp. 1381-5.
- Bornmann, L. and Daniel, H.-D. (2009), “The state of h index research. Is the h index the ideal way to measure research performance?”, *EMBO Reports*, Vol. 10 No. 1, pp. 2-6.
- Braun, T., Glänzel, W. and Schubert, A. (2005), “A Hirsch-type index for journals”, *The Scientist*, Vol. 19 No. 22, p. 8.

- Cronin, B. and Meho, L. (2006), "Using the h-index to rank influential information scientists", *Journal of the American Society for Information Science and Technology*, Vol. 57 No. 9, pp. 1275-8.
- Egghe, L. (2006), "Theory and practise of the g-index", *Scientometrics*, Vol. 69 No. 1, pp. 131-52.
- Glänzel, W. (2006), "On the h-index: a mathematical approach to a new measure of publication activity and citation impact", *Scientometrics*, Vol. 67 No. 2, pp. 315-21.
- 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*, Vol. 102 No. 46, pp. 16569-72.
- Jacso, P. (2008a), "The pros and cons of computing the h-index using Web of Science", *Online Information Review*, Vol. 32 No. 5, pp. 673-88.
- Jacso, P. (2008b), "The pros and cons of computing the h-index using Scopus", *Online Information Review*, Vol. 32 No. 4, pp. 524-35.
- King, D. (2004), "The scientific impact of nations", *Nature*, Vol. 430 No. 6697, pp. 311-6.
- Levitt, J.M. and Thelwall, M. (2009), "The most highly cited library and information science articles: interdisciplinarity, first authors and citation patterns", *Scientometrics*, Vol. 78 No. 1, pp. 45-67.
- Meho, L.I. and Rogers, Y. (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*, Vol. 59 No. 11, pp. 1711-26.
- Meneghini, R. and Packer, A.L. (2006), "Articles with authors affiliated to Brazilian institutions published from 1994 to 2003 with 100 or more citations: II – identification of thematic nuclei of excellence in Brazilian science", *Anais Da Academia Brasileira De Ciencias*, Vol. 78 No. 4, pp. 855-83.
- Molinari, J. and Molinari, A. (2008), "A new methodology for ranking scientific institutions", *Scientometrics*, Vol. 75 No. 1, pp. 163-74.
- Moravcsik, M.J. (1985), "Applied scientometrics: an assessment methodology for developing countries", *Scientometrics*, Vol. 7 No. 3, pp. 165-76.
- Oppenheim, C. (2007), "Using the h-index to rank influential British researchers in information science and librarianship", *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 2, pp. 297-301.
- Prathap, G. (2006), "Hirsch-type indices for ranking institutions' scientific research output", *Current Science*, Vol. 91 No. 11, p. 1439.
- Rousseau, R. (2007), "The influence of missing publications on the Hirsch index", *Journal of Informetrics*, Vol. 1 No. 1, pp. 2-7.
- Rousseau, R. and Ye, F.Y. (2008), "A proposal for a dynamic h-type index", *Journal of the American Society for Information Science and Technology*, Vol. 59 No. 11, pp. 1853-5.
- Schreiber, M. (2007), "Self-citation corrections for the Hirsch-index", *Europhysics Letters*, Vol. 78 No. 3.
- Schubert, A. and Glänzel, W. (2007), "A systematic analysis of Hirsch-type indices for journals", *Journal of Informetrics*, Vol. 1 No. 3, pp. 179-84.
- Vanclay, J.K. (2007), "On the robustness of the h-index", *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 10, pp. 1547-50.

Van Rann, A.F.J. (2006), "Statistical properties of bibliometric indicators: research group indicator distributions and correlations", *Journal of the American Society for Information Science and Technology*, Vol. 57 No. 3, pp. 408-30.

**Further reading**

- Bornmann, L., Mutz, R., Neuhaus, C. and Daniel, H.-D. (2008), "Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results", *Ethics in Science and Environmental Politics*, Vol. 8 No. 1, pp. 93-102.
- Egghe, L. and Rousseau, R. (2008), "An h-index weighted by citation impact", *Information Processing and Management*, Vol. 44 No. 2, pp. 770-80.
- Schreiber, M. (2008), "A modification of the h-index: the hm-index accounts for multi-authored manuscripts", *Journal of Informetrics*, Vol. 2 No. 3, pp. 211-6.