Calculating the h-index: Web of Science, Scopus or Google Scholar?

- The h-index, a method of measuring the productivity and impact of an academic’s work, is often used as a component or metric in the ranking of higher education institutions and their staff.
- The principle citation databases used in this exercise are: Web of Science (WoS), Scopus and Google Scholar (GS) and the pros and cons of using each of the three databases to calculate the h-index are discussed below.
- The h-index defined: a researcher has an h-index of 25 if 25 of their papers have been cited at least 25 times.
- These databases are selective in their journal coverage and some disciplines are better served than others. Also, conference proceedings and monographs, which are key research communication channels in some subjects, are not adequately covered. These factors should be kept in mind when assessing the h-indices of researchers in such disciplines.

### Web of Science

**Pros**
- Excellent depth of coverage in the full product (from 1900-present for some journals).
- A large number of the records are enhanced with cited references.
- Regional journal coverage has been improved.
- The first database to incorporate the h-index (using a good graphical display).
- You can view the h-index minus self citations, though this only removes them for the first author.
- Conference proceedings coverage has improved.
- You can view stray and orphan records using the “cited references” search feature to expand the citation set.

**Cons**
- Coverage of journal titles is not as wide as Scopus (around 10,000 active journals).
- Better coverage of sciences than arts and humanities.
- Does not cover monographs in any depth.
- Facilities for finding and distinguishing between authors are not great in the product.
- Western, US and English language bias.

**Example:**

H-index values* for Stephen Hawking:

<table>
<thead>
<tr>
<th>Database</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>WoS</td>
<td>59</td>
</tr>
<tr>
<td>Scopus</td>
<td>19</td>
</tr>
<tr>
<td>GS</td>
<td>76</td>
</tr>
</tbody>
</table>

*Generated in each database using the automatic h-index tool provided (for GS, Publish or Perish software used). No manual verification conducted.

### Google Scholar

**Pros**
- Covers not only journals but academic websites, grey literature, pre-prints, theses etc.
- Also includes books from the Google Books project.
- Includes electronic-only publications.
- Adds a record to the database for every cited work.
- Adds a record to the database for every cited work.
- Includes more than 18,000 journal titles.
- Covers some books and conference proceedings.
- The “more” feature allows you to quickly view stray and orphan records.
- Very strong coverage of science and technology journals and full Medline coverage.
- Contains useful tools for author disambiguation.
- Automatically generates the h-index.

**Cons**
- Does not automatically calculate the h-index (but you can use “Publish or Perish” software to do this or try the Scholarometer add-on for Firefox).
- Does not provide a list of journals and other items indexed and covered (peer reviewed or otherwise).
- Does not indicate the timescale covered.
- Covers some non-scholarly or research level sources e.g. course reading lists, student projects.
- Poorer coverage of print-only material than rivals.
- Difficult to distinguish between authors who have the same initials in this product, so a lot of manual check and pruning of result sets is needed.
- Results often contain duplicates of the same article (usually as pre-prints and post-prints) due to the wide range of sources.

### Scopus

**Pros**
- Depth of coverage is not as impressive as the width, many journals are only covered for the last 5 years.
- Relatively poor coverage of arts and humanities (recently improved somewhat as more journals added).
- The citations and calculations based on them are only available from publications since 1996. This results in a very skewed h-index for researchers with longer careers than this.
- Citations to pre-1996 articles in articles published after 1996 are not included in the h-index calculation.

**Cons**
- Poorer coverage of print and websites, grey literature, pre-prints, theses etc.
- Does not cover monographs in any depth.
- Facilities for finding and distinguishing between authors are not great in the product.
- Western, US and English language bias.

### References


Website: [http://www.ndlr.ie/myri/](http://www.ndlr.ie/myri/)

Last edited: 17 Apr 2011