# Hirsch-index (2005)

**(Synthesis from Wikipedia, the Free Encyclopedia)**

**The Hirsch index seemns to be the best and most appreciated index for the characterization of the scientific creativity of a researcher or university professor. On its basis it is possible to make a hierarchy of the specialists in a given field of research. The Hirsch index seems to be much better than the classical impact factor (IF).**

The ***h*-index** is an index that quantifies both the actual scientific productivity and the apparent scientific impact of a scientist. The index is based on the set of the scientist's most cited papers and the number of citations that they have received in other people's publications. The index can also be applied to the productivity and impact of a group of scientists, such as a department or university or country. The index was suggested by Jorge E. Hirsch in 2005 as a tool for determining theoretical physicists relative quality[1] and is sometimes called the *Hirsch index* or *Hirsch number*.

Hirsch suggested that, for physicists, a value for h of about 12 might be a useful guideline for tenure decisions at major research universities. A value of about 18 could mean a full professorship, 15–20 could mean a fellowship in the American Physical Society, and 45 or higher could mean membership in the United States National Academy of Sciences[2].

The index is based on the distribution of citations received by a given researcher's publications. Hirsch writes:

> *A scientist has index h if h of his $N_p$ papers have at least h citations each, and the other ($N_p$ - h) papers have at most h citations each.*

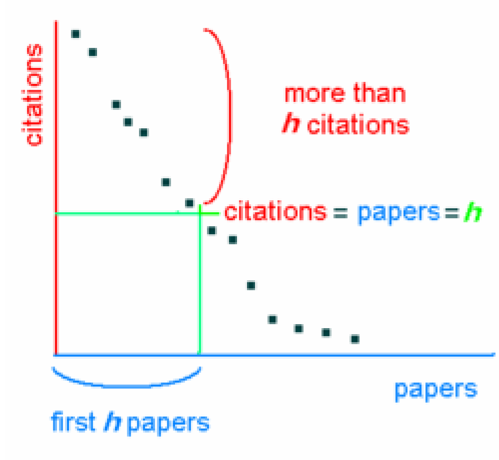


Fig. 1 shows the determination of the H-index from a plot of decreasing citations versus numbered papers.

In other words, a scholar with an index of *h* has published *h* papers each of which has been cited by others at least *h* times.[1] Thus, the *h*-index reflects both the number of publications and the number of citations per publication. The index is designed to improve upon simpler measures such as the total number of citations or publications. The index works properly only for comparing scientists working in the same field; citation conventions differ widely among different fields.

The *h*-index serves as an alternative to more traditional journal impact factor metrics in the evaluation of the impact of the work of a particular researcher. Because only the most highly cited articles contribute to the *h*-index, its determination is a relatively simpler process. Hirsch has demonstrated that *h* has high predictive value for whether a scientist has won honors like National Academy membership or the Nobel Prize. In physics, a moderately productive scientist should have an *h* equal to the number of years of service while biomedical scientists tend to have higher values.

The *h*-index can be manually determined using free Internet databases, such as Google Scholar. Subscription-based databases such as Scopus and the Web of Knowledge provide automated calculators. Each database is likely to produce a different *h* for the same scholar, due to different coverage in each DB: Google Scholar has more citations than Scopus and Web of Science but each of their smaller citation collections tends to be more accurate.

The topic has been studied in some detail by Meho and Yang[3]. Web of Knowledge was found to have strong coverage of journal publications, but poor coverage of high impact conferences (a particular problem for Computer Science based scholars); Scopus has better coverage of conferences, but poor coverage of publications prior to 1992; Google Scholar has the best coverage of conferences and most journals (though not all), but like Scopus has limited coverage of pre-1990 publications. Google Scholar has also been criticized for including gray literature in its citation counts[4]. However, the Meho and Yang study showed that the majority of the additional citation sources Google

Scholar uses are legitimate refereed forums. It has been suggested that in order to deal with the sometimes wide variation in *h* for a single academic measured across the possible citation databases, that one could assume false negatives in the databases are more problematic than false positives and take the maximum *h* measured for an academic[5].

It should be remembered that the content of all of the databases, particularly Google Scholar, continually changes, so any research on the content of the databases risks going out of date.

The *h*-index was intended to address the main disadvantages of other bibliometric indicators, such as total number of papers or total number of citations. Total number of papers does not account for the quality of scientific publications, while total number of citations can be disproportionately affected by participation in a single publication of major influence. The *h*-index is intended to measure simultaneously the quality and sustainability of scientific output, as well as, to some extent, the diversity of scientific research. The *h*-index is much less affected by methodological papers proposing successful new techniques, methods or approximations, which can be extremely highly cited. For example, one of the most cited condensed matter theorists, John P. Perdew, has been very successful in devising new approximations within the widely used density functional theory. He has published 3 papers cited more than 5000 times and 2 cited more than 4000 times. Several thousand papers utilizing the density functional theory are published every year, most of them citing at least one paper of J.P. Perdew. His total citation index is close to 39 000, while his *h*-index is large, 51, but not unique. In contrast, the condensed-matter theorist with the highest *h*-index (94), Marvin L. Cohen, has a lower citation index of 35 000. One can argue that in this case the *h*-index reflects the broader impact of Cohen's papers in solid-state physics due to his larger number of highly-cited papers.

There are a number of situations in which *h* may provide misleading information about a scientist's output[6]

- The *h*-index is bounded by the total number of publications. This means that scientists with a short career are at an inherent disadvantage, regardless of the importance of their discoveries. For example, Évariste Galois' *h*-index is 2, and will remain so forever. Had Albert Einstein died in early 1906, his *h*-index would be stuck at 4 or 5, despite his being widely acknowledged as one of the most important physicists, even considering only his publications to that date.

- The *h*-index does not consider the *context* of citations. For example, citations in a paper are often made simply to flesh-out an introduction, otherwise having no other significance to the work. *h* also does not resolve other contextual instances: citations made in a negative context and citations made to fraudulent or retracted work. (This is true for other metrics using citations, not just for the h-index.)

- The *h*-index does not account for confounding factors. These include the practice of "gratuitous authorship", which is still common in some research cultures, the so-called Matthew effect, and the favorable citation bias associated with review articles.

- The *h*-index has been found to have slightly less predictive accuracy and precision than the simpler measure of mean citations per paper.[7]

- While the *h*-index de-emphasizes singular successful publications in favor of sustained productivity, it may do so too strongly. Two scientists may have the same *h*-index, say, $h = 30$, but one has 20 papers that have been cited more than 1000 times and the other has none. Clearly scientific output of the former is more valuable. Several recipes to correct for that have been proposed, such as the g-index, but none has gained universal support.

- The *h*-index is affected by limitations in citation data bases. Some automated searching processes find citations to papers going back many years, while others find only recent papers or citations. This issue is less important for those whose publication record started after automated indexing began around 1990. Citation data bases contain some citations that are not quite correct and therefore will not properly match to the correct paper or author.

- The *h*-index does not account for the number of authors of a paper. If the impact of a paper is the number of citations it receives, it might be logical to divide that impact by the number of authors involved. (Some authors will have contributed more than others, but in the absence of information on contributions, the simplest assumption is to divide credit equally.) Not taking into account the number of authors could allow gaming the *h*-index and other similar indices: for example, two equally capable researchers could agree to share authorship on all their papers, thus increasing each of their *h*-indices. Even in the absence of such explicit gaming, the *h*-index and similar indices tend to favor fields with larger groups, e.g. experimental over theoretical.

The *h*-index grows as citations accumulate and thus it depends on the 'academic age' of a researcher. Using papers published within a particular time period, e.g within the last 10 years, would allow to measure the current productivity as opposed to the lifetime achievement.

# References


1. Hirsch, J. E. (2005). "An index to quantify an individual's scientific research output," *Proceedings of the National Academy of Sciences*, **102**(46):16569-16572, November 15, 2005 (Free copy available from arXiv).
2. Math Trek: Rating Researchers, Science News Online, Dec. 3, 2005
3. Meho, L.I., Yang, K. (2007) "Impact of Data Sources on Citation Counts and Rankings of LIS Faculty: Web of Science vs. Scopus and Google Scholar" to

appear in Journal of the American Society for Information Science and Technology

4. Jacsó, Péter (2006) "Dubious hit counts and cuckoo's eggs" Online Information Review Volume 30 Number 2 2006 pp. 188-193
5. Sanderson, Mark (2008) "Revisiting *h* measured on UK LIS academics," to appear in Journal of the American Society for Information Science and Technology
6. Wendl, Michael (2007) "H-index: however ranked, citations need context," Nature **449**(7161):403.
7. Sune Lehmann, Andrew D. Jackson, Benny E. Lautrup (2006). "Measures for measures". *Nature* **444** (7122): 1003–4. doi:10.1038/4441003a. PMID 17183295.
8. Sidiropoulos A., Katsaros D. and Manolopoulos Y., (2006), Generalized h-index for disclosing latent facts in citation networks
9. g-index
10. Jayant S Vaidya (December 2005). "V-index: A fairer index to quantify an individual 's research output capacity". *BMJ* **331**: 339-c-1340-c.
11. Katsaros D., Sidiropoulos A., Manolopous Y., (2007), Age Decaying H-Index for Social Network of Citations in Proceedings of Workshop on Social Aspects of the Web Poznan, Poland, April 27, 2007