

The h Index in Science: A New Measure of Scholarly Contribution

April 01, 2006

Debates have long swirled about how to measure the contributions of scientists and other scholars. Such assessments are necessary for hiring, for promotion and tenure, for raises, for election to honorary societies, and for awards, among other purposes. Scholarly productivity is one such measure. It is easy enough to count publications and also relatively easy to parcel them into various categories: refereed journal articles, book chapters, books, and so on. Further, one can judge the quality of the journal for articles (perhaps using impact factors) and the quality of the books published (a major publisher or more nearly a vanity press). However, we all know that scholars can be quite productive, publishing even in the best of journals, without their work having much influence on the field. The impressive number of publications amassed by some researchers may represent much ado about not very much, the bane of all scholarly publishing. So counting publications does not necessarily tell much, except for the obvious point that in order to publish something important and influential, it is first necessary to publishing something.

Enter the citation index. The Institute of Scientific Information has been faithfully assessing citations to scholarly publications for many years by indexing the reference lists of articles published in journals. As long as a person does not have a common name, it is fairly easy to determine that person's citation counts. (It is possible to get accurate counts for the John Smiths of the world, too, but usually one needs the particular John Smith's vita). Citations measure the impact of one's work, the frequency with which other people are using it. A high citation count means that someone has great impact on the field in the sense that many other researchers need to cite that person's work in their own work. A paper with a high citation count has, *ipso facto*, high impact. Yes, interpretive problems exist in using citations, too. What about self-citations? For papers or people with very high counts, they aren't much of a problem. What about "negative citations," or citations to people's work because it is shoddy or wrong? I'm not really convinced this problem exists; work that is really bad is usually ignored after a paper or two discredits it, and if a paper ignites a debate that persists, the topic is usually really important (even if the initial paper attracted critics).

Still, if a person has a very high count of total citations, it does not necessarily mean that he or she has a large body of excellent work, because citations are a notoriously skewed measure. When Endler, Rushton, and Roediger published an analysis of 1975 citation data for psychology departments and psychologists (in the *American Psychologist*, 1978), the most cited living psychologist for that year was Ben Winer of Purdue University. Virtually every citation was to his then widely used textbook, *Statistical Principles in Experimental Design* (McGraw-Hill, 1962). Winer actually had relatively few publications to his credit, but he had great impact with one book, which was widely used at the time. Some might argue that total citations as a measure of faculty quality may leave something to be desired because the citations could be based on one or two very influential publications.

Is there a way to assess simultaneously both publications and citations, to provide an index of the

number of highimpact contributions that a person publishes? Various methods exist. For example, a department chair could assess his or her department members by determining how many papers each person has published that are cited more than 25 times, 100 times, or 250 times. This is perfectly valid, but takes a fair amount of research. Is there a better, simpler way? Is there a single number that might capture both publications and their impact?

Yes, there is now. Jorge E. Hirsch, a physicist at the University of California at San Diego, has devised a new measure that he eponymously calls h . The beauty of h is that a single number represents both productivity and citation impact in one simple measure; h is defined as the number of papers that a person has with citations equal to or greater than h . So, if Harry Franzdorf has an h of 10, then Harry has 10 papers with at least 10 citations. All his other papers, however many there may be, have fewer than 10 citations. Hirsch's paper proposing h (published in *Proceedings of the National Academy of Sciences* in 2005 and easily available via Google — just type in “ h index”) notes many advantages of h relative to other measures (total publications or citations, average number of citations per paper).

One advantage of h is that it can be easily obtained by anyone with access to the Institute of Scientific Information Web of Knowledge. Here is how to find h for a given person. I'll use B.F. Skinner here as an example of the method (although the number obtained underestimates Skinner's h , for reasons I'll describe later). First, get on the ISI Web of Knowledge and click on Web of Science. For senior researchers, it is necessary to access the “expanded version” in which the Science Citation Index extends back to 1945 (which is still not far enough for Skinner, who began publishing in the 1930s). Make sure the settings are for “all years” which is usually the default setting. Second, find and click on “General Reference Search.” You will now get a screen listing such items as “Topic” and “Author.” If you want to find a person's h , type his or her name into the Author box in the prescribed form (Skinner, BF*). Then press “Search.” After a few seconds, the screen will show all the papers by B.F. Skinner represented in the ISI data base (149) and sorted from the most recent article. Now find a window on the right that says “Sort By” and click on the “Times Cited” feature. In a few more seconds, all Skinner's papers will be ordered by the number of times they were cited, with the most cited listed first. In his case, “Are theories of learning necessary?” in *Psychological Review* is first, having been cited 580 times. The final step in calculating h is counting down until the number of papers equals the number of citations. In Skinner's case, his 32nd listed paper was cited 32 times, so Skinner's h is 32. (Even though he is deceased, his h can continue to climb as his work is cited, so there is h after death).

In an article by Adrian Cho in *Science* magazine (12 August, 2005) entitled “Your Career in a Number,” he commented that “The h index favors researchers who consistently produce influential papers and disfavors those who publish many little-noted papers or just a few highly cited ones.” In the same article Hirsch was quoted as saying “I can't imagine a person with a high h index who hasn't done important work.” Hirsch has examined h for physicists who have won the Nobel Prize in the past 20 years and another group that had been recently elected to the National Academy of Sciences. The mean of the h distribution for Nobel Prize winners was 41 with a standard deviation of 15. The median was 35 and the range from 22 to 79. However, the highest physicist he found was Edward Witten at the Institute for Advanced Study at Princeton University, with an h of 110 (110 papers cited 110 or more times).

Hirsch noted in his paper that fields would probably differ widely in their range of h . He checked some leading researchers in the biological sciences and found h s much higher than those of the top physicists. For example, David Baltimore, the president of the California Institute of Technology, has an h of 160.

According to Hirsch, h provides “an estimate of the importance, significance and broad impact of a scientist’s cumulative research and contributions. I suggest that this index may provide a useful yardstick to compare different individuals competing for the same resource when an important evaluation criterion is scientific achievement, in an unbiased way.”

That last statement is a strong claim, because of course Hirsch’s measure does have biases built into it. He was aware of many of these and wrote about them. Take some obvious ones: h will be generally correlated with age of the researcher, because the measure can only go in one direction, up. (As just noted, the person need not be alive to have an increasing h). Another measure, m , which essentially measures the slope of the line relating h to years in the field (in most cases, the years since earning a PhD) can be used as a normalization factor for age.

Lag is a problem, too. In theory, a scientist could have published 40 papers between 1980 and 1990 and then stopped publishing because he or she left the field and went into business. This person’s h might have been zero in 1990, but then in 2006 (after 16 years of scientific inactivity), his or her h could be 35 if others had found and cited the work. Of course, one can look for immediacy or recency of papers contributing to h , but a bias is built in here, too. If David Baltimore publishes a paper in 2006, it has to be cited over 160 times before it can contribute to his h .

To my knowledge, there has been no thorough study of psychological scientists using h as the measure, and I am not going to provide one here. However, I will offer some observations based on my rather haphazard survey of a few people and based on my assistant, Jane McConnell, having looked up full professors in my own department. The mean h of 13 of our 14 full professors (excluding one due to his having a common name and initial that prevented my easily obtaining his h) is 30.6 (the median is 28) with a standard deviation of 14.26. The range is from 9 to 64. The 64 belongs to Endel Tulving, who has been the Clark Way Harrison Distinguished Visiting Professor in our department for two to three months each year for the past 10 years. Without Tulving, the mean, standard deviation, and median and mode are 28.1, 11.0, and 28, respectively. I checked just a couple of other faculty members who I thought might be fairly highly cited and found an associate professor with an h of 24 and an assistant professor (about to be promoted) with an h of 21.

In performing this quick exercise, I also discovered some other features of h that are troubling, interesting, or both. First, at least as far as I can tell, the customary method used in calculation of h (as recommended above) usually leads to a wrong number for senior members of the academy. h underestimates a researcher’s influence. Now, I know some of you nitpickers will start losing interest here, but the fact that h is often wrong does not mean that there is anything wrong with the measure per se. The problem arises because of the peculiar characteristics of the ISI Web of Science. The three databases that go into the ISI (the Science Citation Index, dating back to 1945; the Social Science Citation Index, dating back to 1956; and the Arts and Humanities Index, dating back to 1975) use citations published in journal articles. Singleauthored books and most edited books are not included, so citations appearing in those works are not counted. Most people using the ISI understand this characteristic, and it would be incredibly difficult to list references in all books. However, a consequence is that, unlike journal articles, these books and chapters are not listed in the source index and do not come up in the “General Reference Search” used to calculate h . Citations (from journal articles) to books and book chapters do exist in the database, but they are not called up in the General Reference Search. One must find them through a “Cited Reference Search” in which an author and year are typed in (if the

year is known) or the author and title of the book or chapter.

Ignoring books and book chapters might work in some scientific fields, but in psychology many important contributions still occur in books and book chapters, which can be highly cited. Tulving's *Elements of Episodic Memory* (Oxford, 1983) has been cited 1,718 times, but does not count towards his h . Similarly, his 1972 chapter on "Episodic and semantic memory" has been cited 1,577 times and yet does not contribute to his h . If chapters and books are added to the total, Tulving's h goes from 64 to 70 (which represents a huge jump, because the higher one's h , the harder it is to increase it). By the way, Skinner's *Science and Human Behavior* (1953) has been cited 3,200 times but does not contribute to his h , either. Skinner's total impact on psychology is much greater than his h of 32 would have one believe, because h counted only journal articles and the various indices only began tracking articles long after Skinner began publishing in the 1930s. His h , could we accurately measure it, would be much higher due to the earlier articles and his books. Also, it is probably difficult to compare h of people across time because of changes in interests in the field.

If we calculate a person's true h index (that is, including citations to books and book chapters), the figure might be given some other notation, perhaps h . I did some spot checking in my department, and most senior faculty members' numbers rise a bit when books and book chapters are included (but actually not as much as I expected). In my case, I go from an h of 39 to an h of 43 thanks to some heavily cited chapters. Seven chapters are included in the 43, but they bumped out three papers cited between 39 and 42 times, so the total did not go up by 7.

There are two ways to calculate h . One is to have a person's vita and to look up the books and book chapters with a Cited Reference Search in the ISI, then compute H by taking the h data and appropriately putting in figures for chapters and books. (As just noted above, the numbers don't necessarily just go up as a function of chapters and books cited more than h times, because some articles may drop off). If one does not have access to the person's vita, one can perform a Cited Reference Search for each year since the person began publishing and look for highly cited chapters, but this would be time-consuming. Because the citation data for books and chapters all appear in the ISI, it seems a shame that the ISI/Web of Science database cannot be altered so that when one conducts a General Reference Search, all cited entries in the database are consulted. (To reiterate: As it is now, only journal publications and certain types of book chapters, in books such as the *Annual Review of Psychology*, are identified.)

Hirsch extols the advantage of h not depending on total citations, which is true (and true of H , too). However, shouldn't total citations count? Aren't some papers truly great advances that are highly cited for the intellectual achievement they represent? For example, I have discovered people with rather similar h values but with hugely different impact. Imagine two faculty members who both have an h of 25. However, one person's top five cited papers may have 125, 110, 95, 84, and 70 citations (which is quite good), whereas for the other person the corresponding numbers could be 1,650, 1,508, 1,433, 1,226, and 947 (much, much higher). The general point is that h has a property similar to the mode, in that it is not sensitive to extreme values. Yet in science it is the extremely valuable and influential papers that usually represent the great breakthroughs, and computing only h can miss this feature. Perhaps a person's career cannot be captured by a single number (although probably h and total citations would be highly correlated if anyone did the appropriate study).

Another issue that arises is co-authorship and how to apportion credit. h is benign, and everyone gets

credit. However, in examining h for some neuroscientists, some quirks appear. This is a field that tends to have many coauthors and usually the first author and last author are considered as most responsible and most deserving of credit (with others having contributed their justifiable, but lesser, pieces to the project). However, if the paper goes on to be highly cited, the paper contributes to the h of all individuals on the paper, whether there were three authors or 15 authors. Some neuroscientists with high values of h may achieve them in part via dozens of papers on which they are from 2nd to 6th author (with, say, 8 authors on the paper). In such cases of huge lists of authors, should we consider only papers in which authors were first or last? Or is this just one of those differences among fields that we may as well get used to, so that we should not compare h of a cognitive neuroscientist (with large numbers of co-authors) with a cognitive psychologist (typically with few co-authors), despite other similarities in the fields? Hirsch points out that an even more extreme form of this problem exists in high-energy physics (there can be 50 or more co-authors) and suggests that some normalization factor might be considered to ameliorate this problem. For example, one could divide h for individuals by the average number of authors on their papers. If a person has an h of 30 and an average of 5 co-authors, then the person's corrected h would be 6. (Other corrections are possible, of course).

Another issue is recency of citation. I looked up someone whom I knew to have stopped publishing quite a while ago, while still being a full-time faculty member, and was surprised at the person's relatively high h . Sure enough, on checking, all the papers contributing to his h were ones published in the 1970s or before. One solution might be to ask what proportion of papers going into h were contributed in the past 10 years; but as noted above, such a calculation would actually work against people with high h s, even if they remained productive, because each new paper would have to be cited a large number of times to contribute to that person's h .

Do these issues and others (e.g., self-citation issues) mean that h is not a worthwhile new measure? It is too early to say, but my guess is that h is quite valuable. Yes, like virtually all measures of any human quality, h will be shown to have some biases (and various corrections might be undertaken). However, to go out on a limb, I will bet that h measured in the straightforward manner described above will probably correlate highly with H , with total citations, and with various other corrections. Because it provides a simultaneous measure of both publications and their impact, h seems a useful scientific measure that is probably here to stay. Certainly more careful research will be needed to document this claim than my own cursory explorations in this essay, but I can hope others will be inspired to take up the challenge. Baseball has an army of sabermetricians, as they are called, who measure players from every conceivable angle (e.g., batting average for a player batting with a man on third, with two outs, against a left-handed pitcher). One could wish that psychology would develop its own army of scientometricians to examine fascinating data on citations and publications.

Jane McConnell assisted in collecting data for this column.