Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht

Scientometrics, Vol. 76, No. 3 (2008) 577–588
DOI: 10.1007/s11192-007-2071-2

Beyond the Durfee square: Enhancing the h-index to score total publication output

THOMAS R. ANDERSON, ROBIN K. S. HANKIN, PETER D. KILLWORTH[†]

National Oceanography Centre, University of Southampton, Southampton (United Kingdom)

An individual's *h*-index corresponds to the number *h* of his/her papers that each has at least *h* citations. When the citation count of an article exceeds *h*, however, as is the case for the hundreds or even thousands of citations that accompany the most highly cited papers, no additional credit is given (these citations falling outside the so-called "Durfee square"). We propose a new bibliometric index, the "tapered *h*-index" (h_T), that positively enumerates all citations, yet scoring them on an equitable basis with *h*.

The career progression of h_T and *h* are compared for six eminent scientists in contrasting fields. Calculated h_T for year 2006 ranged between 44.32 and 72.03, with a corresponding range in *h* of 26 to 44. We argue that the h_T-index is superior to *h*, both theoretically (it scores all citations), and because it shows smooth increases from year to year as compared with the irregular jumps seen in *h*. Conversely, the original *h*-index has the benefit of being conceptually easy to visualise. Qualitatively, the two indices show remarkable similarity (they are closely correlated), such that either can be applied with confidence.

[†] Sadly, after a long and distinguished career, Peter Killworth died on 28 Jan 2008.

Received January 23, 2008

Address for correspondence:
THOMAS R. ANDERSON
National Oceanography Centre, University of Southampton, European Way
Southampton SO14 3ZH, United Kingdom
E-mail: tra@noc.soton.ac.uk

0138–9130/US $ 20.00

Introduction

Numerical indices that permit quantification of published research output are being increasingly used by employers, promotion panels and funding agencies. The *h*-index has risen to the fore since being proposed by HIRSCH [2005]. According to this index, an author scores *h* if *h* of their *N* papers each have at least *h* citations, with the remaining ($N - h$) papers each having fewer than *h* citations. HIRSCH [2005] argued that the *h*-index is superior to other bibliometric indices such as the total number of citations or total papers because it favours authors who consistently produce influential papers rather than those who either publish the occasional highly cited article or many papers that have little or no impact. He further suggested that a scientist might be described as "successful" when achieving an *h*-index of 20 after 20 years, or "outstanding" when scoring *h* of 40 over the same period [HIRSCH, 2005].

An author's *h*-index cannot exceed his/her number of publications, and will usually be considerably less. Thus, the vast majority of the hundreds or even thousands of citations that accompany the most highly cited papers effectively contribute zero, these papers each scoring 1 towards the *h*-index score. Moreover, articles that have received many citations, but which fall just short of the number required to score for *h* (called "sleeping beauties" by [VAN RAAN, 2004]), also count for nothing in the sense that *h* is not affected by them.

We suggest that a bibliometric measure of publication output should be "strictly monotonic", i.e., assign a positive score to each new citation as it occurs. At the very least, outstanding articles with numerous citations should increase the index. EGGHE [2006], for example, proposed an alternative measure, the *g*-index, in which after ranking papers in order of decreasing citations, *g* is the (unique) largest number that together receives g^2 or more citations. Along similar lines, JIN [2006] defined an index based on the average number of citations received by those articles that contribute to the *h*-score.

Here, we describe a new version of the *h*-index, the "tapered *h*-index" (h_T), which positively scores all of an author's citations (i.e., it is strictly monotonic), accounting for the tapered distribution of citations associated with highly cited papers rather than using a cut-off at *h*. An advantage of our approach is that the scoring mechanism of h_T is on an equitable basis to that of *h*, permitting direct comparison of the two measures of output. Calculations are presented for each index, firstly theoretical examples, and then an analysis of how each is reflected in the career progress of six eminent scientists. The merits of the two indices are discussed.

Tapered *h*-index

Consider a scientist who has 5 publications which, when ranked, have 6,4,4,2,1 citations. This publication output can be represented by a Ferrers graph, where each row represents a partition of the total 17 cites amongst papers (Figure 1). The largest completed (filled in) square of points in the upper left hand corner of a Ferrers graph is called the Durfee square (e.g., [ANDREWS, 1984]). The *h*-index is equal to the length of the side of the Durfee square (in the case of Figure 1, $h = 3$), effectively assigning no credit (zero score) to all points that fall outside.

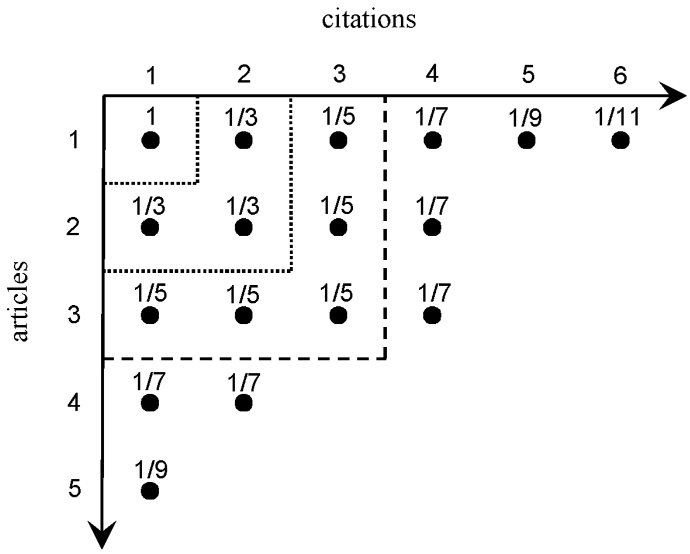

Figure 1. Example of a Ferrers diagram of an author's citations, in this case with 5 papers and a total of 6+4+4+2+1 = 17 citations, indicated in rows. The Durfee square is the 3-by-3 square indicated by a dashed line; this is the largest complete square in the Ferrers diagram [ANDREWS' 1984]. For convenience, any complete square of side *L* with a corner at the origin is referred to as "the Durfee square of side *L*". Citation scores are shown according to the tapered *h*-index, h_T (see text).

Let us start by considering *h*-index scores for sets of citation records that exactly match Durfee squares. If an author has a single paper that has one citation, this scores $h = 1$. Subsequently, $h = 2$ is achieved with two papers each with two citations. To move from $h = 1$ to $h = 2$, an additional 3 citations are required, one for the first paper and two for the second paper. In turn, moving from $h = 2$ to $h = 3$ requires a further 5 citations, reaching a 3, 3, 3 partitioning of the nine citations in the Ferrers graph (and so a Durfee square of side 3). Following this scheme, it is possible to score each citation individually, and in a manner that generates identical *h*-index scores when the relevant Durfee squares are complete (Figure 1). Thus, the single citation in the Durfee square of side one has a score of 1, the three additional citations in the Durfee square of side 2 each score 1/3, and the five additional citations in the Durfee square of side 3 each score 1/5. Summing the relevant citations, scores of 1, 2, 3 are achieved for Durfee squares whose width is 1, 2, 3, matching the *h*-index.

This notation immediately suggests a new index, h_T, which has the property that each additional citation increases the total score (the index has the property of being "marginally increasing"), whether or not it lies within the *h*-index Durfee square. The score of *any* citation on a Ferrers graph is now given by $1/(2L - 1)$, where L is the length of side of a Durfee square whose boundary includes the citation in question. The additional citations that fall outside the Durfee square (of side 3) in Figure 1 can now be scored, the five papers achieving scores of 1.88, 1.01, 0.74, 0.29 and 0.11, leading to a total score for h_T of 4.03.

In mathematical terms, the most cited paper in a given list, with n_1 citations, generates a score, $h_{T(1)}$, of:

$$h_{T(1)} = \sum_{i=1}^{n_1} \frac{1}{2i-1} = \ln(n_1)/2 + o(1), \tag{1}$$

where $\ln(n_1)$ is the (natural) log of n_1, and o(1) is mathematical shorthand for a term that approaches zero as n_1 approaches infinity. The resulting score is 2.13 for 10 citations, 3.28 for 100 citations, 4.44 for 1000 citations and 5.59 for 10000 citations (Figure 2). These scores are markedly higher than the score of 1 that the top-ranked paper would score for the *h*-index, increasing asymptotically in proportion to $\log(n_1)$. The paper ranked second in the list scores 1/3 for its first citation, and then 1/3, 1/5, 1/7 etc., for further citations as for the top-ranked paper. Now, if an author has N papers with associated citations $n_1, n_2, n_3, \ldots, n_N$ (ranked in descending order as in a Ferrers graph), the h_T score for any single paper ranked j in the list (with n_j citations), $h_{T(j)}$, is:

$$h_{T(j)} = \frac{n_j}{2j-1}, \; n_j \le j \tag{2}$$

$$h_{T(j)} = \frac{j}{2j-1} + \sum_{i=j+1}^{n_j} \frac{1}{2i-1}, \; n_j > j \tag{3}$$

The total tapered *h*-index for a citation-ranked list of publications, h_T, is then calculated by summing over all the papers in the list:

$$h_T = \sum_{j=1}^{N} h_{T(j)} \tag{4}$$

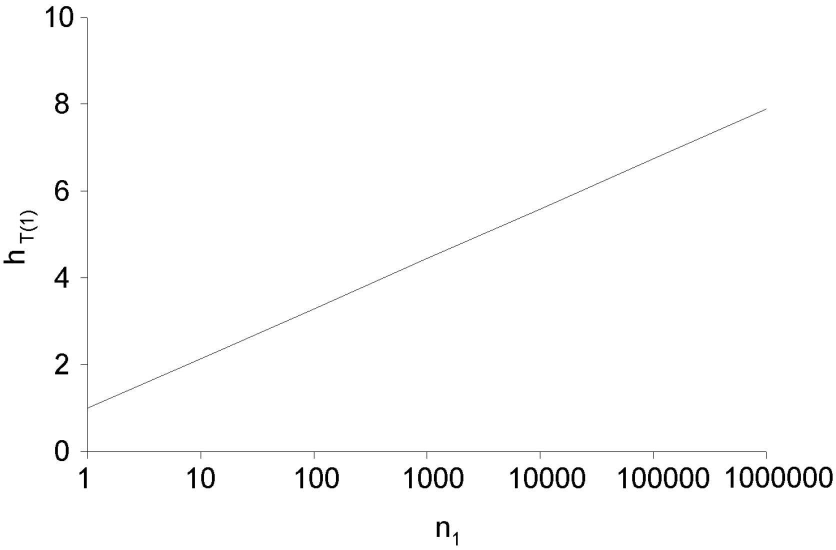

Figure 2. Tapered *h*-index score for an author's top-ranked paper, $h_{T(1)}$, as a function of number of citations (n_1)

Theoretical examples

A series of theoretical examples was provided by VINKLER [2007] that illustrate the advantages and disadvantages of using the *h*-index as a measure of publication performance. Calculated values of *h* and h_T are compared for these examples in Table 1, illustrating the potential benefits of h_T as a bibliometric index. As VINKLER [2007] emphasised, an individual's *h*-index cannot be higher than their total number of papers, putting an immediate restriction on the ability to give additional credit to authors publishing articles of outstanding impact. Compare, for example, authors E and F. Despite having 500 citations, author F scores $h = 5$, which is the same as author E who has a total citation count of just 35. If index h_T is applied instead, however, author F's score is 12.46, which is more than double that of author E (5.79). Similarly, whereas both authors A and B also score $h = 5$, author A scores $h_T = 13.27$, which is approximately double author B's score (6.89) because of five highly cited articles. Author C, with exactly 10 papers each with 10 citations, scores 10 for both *h* and h_T, illustrating the parity between the two indices. Finally, whereas author B also has 10 papers and so scores the maximum possible $h = 10$, his/her h_T index is 18.05 because each has 50 citations, once again accruing considerable extra credit for the large number of additional citations relative to author C.

From this analysis, we conclude that h_T is a fairer index than *h* because it gives due credit to authors who publish leading articles that achieve high citation status.

Table 1. Theoretical examples [VINKLER, 2007]: sets of papers published by authors A, B, etc.
P: total papers, C: total citations; calculated h- and h_T-index are compared

Paper no.	A	B	C	D	E	F
1	100	9	10	50	9	120
2	98	8	10	50	8	110
3	98	8	10	50	7	100
4	97	6	10	50	6	90
5	96	5	10	50	5	80
6	4	4	10	50		
7	3	4	10	50		
8	2	3	10	50		
9	1	2	10	50		
10	**1**	**1**	**10**	**50**		
P	10	10	10	10	5	5
C	500	50	100	500	35	500
h	5	5	10	10	5	5
$\boldsymbol{h_T}$	**13.27**	**6.89**	**10**	**18.05**	**5.79**	**12.46**

Career progression: h and h_T of six eminent scientists

The career progression of h and h_T were determined for six scientists elected to membership of the Royal Society in 2006, chosen at random. We stress that this exercise is in no way about comparing the relative performance of the scientists in question – all are eminent professionals at the top of their representative fields. Intrinsic differences will inevitably arise simply from variations in citation statistics relating to the range of disciplines involved, greater overall citation rates being associated with some disciplines and not others. The aim is rather to compare trends in h and h_T, over the careers of the chosen scientists, in order to assess the suitability of the two indices as measures of publication performance.

Publication output from each scientist was extracted from the Thomson ISI Web of Knowledge database (http://wok.mimas.ac.uk). The "Citation Report" option was used, which generates lists of publications for a chosen author along with the number of times each publication was cited in each year, dating back to 1970. Care was exercised to ensure that errors did not creep into the publication lists by: (1) checking author addresses to exclude articles published by other authors with the same name, and (2) searching variants of author names, including both a full list of author initials and the single first initial, the latter option being preferred by some journals.

Summary statistics for the six scientists, calculated for their careers up to and including the year 2006, are presented in Table 2. The variability between individuals is marked in terms of numbers of papers (P) and total citations (C). Becke, a theoretical chemist, has the fewest publications, 44, but with a remarkable 40,094 citations.

Table 2. Output measures for six randomly chosen fellows of the Royal Society, calculated for their careers up to 2006. Y: year of first citation (i.e. first record included in database); P: total papers (scoring at least one citation); *C*: total citations; h and h_T

Name	Field of expertise	*Y*	*P*	*C*	h	h_T
Barford, D	cancer research	1986	78	6281	41	67.88
Becke, AD	chemistry	1978	55	40094	35	68.18
Lockwood, M	astrophysics	1981	176	5101	39	65.43
Jackson, RJ	molecular biology	1970	79	10778	44	72.03
Proctor, MRE	applied mathematics	1975	89	2356	26	44.32
Saibil, HR	molecular biology	1976	80	4234	33	55.78

Conversely, Lockwood, an astrophysicist, has the most papers, 176, but with 5101 citations. Given this large range of variability in *P* and *C*, neither would seem to be a reliable index of publication impact. In contrast to numbers of papers and citations, variability in h and h_T is much lower, 26–44 and 44.32–72.03 respectively.

The progression of h and h_T throughout the careers (until 2006) of the six scientists, so-called *h*-sequences [LIANG, 2006], are shown in Figure 3. The resulting trends in h and h_T are qualitatively remarkably similar, although much smoother in the case of h_T compared to the somewhat irregular h jumps from one integer score to the next. Performance, as reflected by rates of increase of h and h_T, was also similar between the chosen individuals (typically ~1–2 per year for h, and ~2–4 per year for h_T), at least during the middle to latter stages of the career paths as shown in Figure 3. Early on, however, rates of increase were often lower, presumably as the scientists sought to become established in their field. While inconclusive with such a small sample of scientists, the results nevertheless suggest that there are differences in h and h_T related to discipline rather than performance *per se*. The rate of increase of both indices in the case of Barford (cancer research) is approximately double that of Proctor (applied mathematics). In general, it appears that the score for Proctor is lower than his compatriots in the fields of physics, chemistry and biology.

A plot of h_T versus h is shown in Figure 4, illustrating the close correlation between the two indices. Using a least-squares straight line fit, constrained to pass through (1,1), the ratio of h_T to h is 1.73 ± 0.1; using the individual (maximum likelihood) variances, on the assumption of independent residuals, allows one to reject the null hypothesis that the lines are all of identical slope ($p < 10^{-9}$). The small p-value is due to the goodness-of-fit of the linear relationships for the data relating to each author shown in Figure 4, rather than the (small) difference between the slopes.

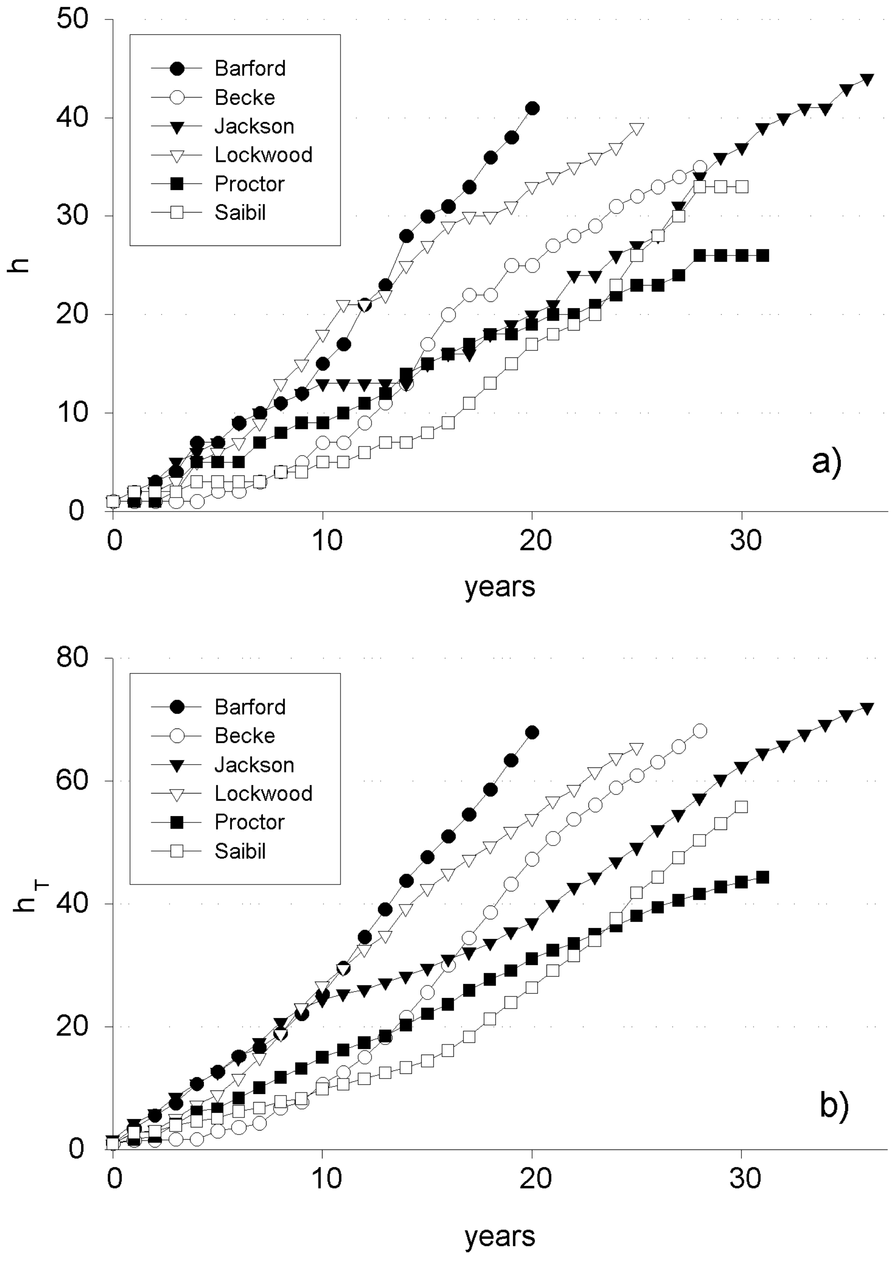

Figure 3. Calculated career progression (until 2006) of: a) h and b) $h_{T,}$ for six eminent scientists (Table 2). Data are normalised to the first year in which a citation was recorded (Y) in order to facilitate the comparison

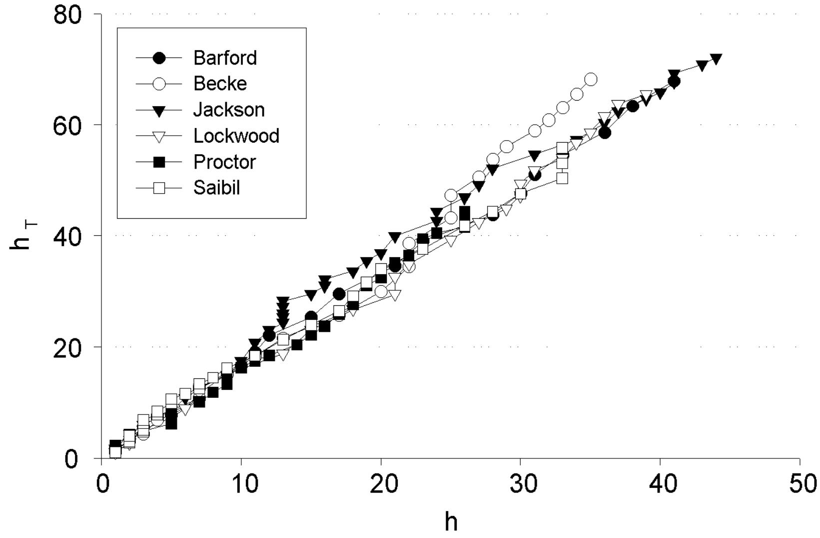

Figure 4. Plot of h_T versus h during the careers of six eminent scientists (data as in Figure 3)

Discussion

It is well known that the distribution of citations among published articles is strongly skewed. In a study of three biochemical journals, SEGLEN [1992] found that the top-ranked 15% of articles accounted for 50% of the total citation count. Similar skewed citation distributions were found by OPTHOF & AL. [2004] and REDNER [2005] when studying citation statistics in the fields of cardiovascular research and physics, respectively. Highly-cited papers are usually regarded as those representing excellent scientific research, in which case bibliometric measures of publication output should score them accordingly. This train of thought was used when developing the impact factor, which measures the frequency with which the articles published in a chosen journal are cited in the scientific literature. Resulting scores are generally considered to be a reasonable indicator of the quality of different journals [SAHA & AL., 2003], although caution should be exercised when extending such analyses to the study of citation counts for individual researchers [OPTHOF, 1997] due to the much smaller sample size.

The *h*-index proposed by HIRSCH [2005] took the information science community by storm because it provides a scoring mechanism based on citation count, but which requires ongoing consistent publication of highly cited articles in order to progress one's score, rather than the production of occasional citation classics. The index has been promoted in *Nature* [BALL, 2005] and *Science* [HOLDEN, 2005], as well as receiving acclaim within the information science literature [BORNMANN & DANIEL, 2005; CRONIN & MEHO, 2006; OPPENHEIM, 2007]. Potential deficiencies in the approach have nevertheless been identified. The skewed nature of the distribution of citations among publications may often result in scientists having several papers that nearly but not quite count when scoring h. These so-called "sleeping beauties" [VAN RAAN, 2004] spring to life when they achieve enough citations to be counted in the h score. Perhaps the greatest criticism that can be levelled at the h index is that an individual's score cannot exceed their number of papers (and will usually be much less), in which case the thousands of citations that typically accompany leading articles in a given field are left effectively unscored (in practice, most of them fall to the right of the Durfee square).

Our new tapered h-index, h_T, has the advantage that it is marginally increasing, and moreover uses a scoring method that is consistent with the original h-index as proposed by HIRSCH [2005]. In this way, due credit is assured for authors who publish outstanding articles that achieve hundreds or thousands of citations. As such, we argue that h_T is a more useful way of scoring publication output than h, at least from a theoretical perspective. The efficacy of the new index was assessed by examining the career progression of h_T for six Fellows of the Royal Society in contrasting fields of study. Smooth and progressive increases are seen over time, which are often approximately linear over periods of many years. When at their most productive, these eminent individuals typically had increases in h_T of between 2 and 4 per year. Different rates of publication output can be expected in different fields [HIRSCH, 2005; VINKLER, 2007]; there was indeed considerable variability in career progression of h_T between the individuals in our study. Proctor, an applied mathematician, scored more slowly than other researchers in physics, chemistry and biology. The outstanding achievement of all the scientists examined is not in question, any differences in individual's scores being most likely a result of overall citation rates in their respective fields.

Suggesting that h increases approximately linearly with time, HIRSCH [2005] defined parameter m as an individual's h value divided by their number of years of service. "Successful" scientists, he proposed, achieve an h-index of 20 after 20 years ($m = 1$), while "outstanding" scientists amass an h of 40 over the same period ($m = 2$). The trends seen in h and h_T of the six scientists we studied here were often linear over periods of years, but generally not so over time spans as long as 20 years. Rates of increase of h and h_T were in some instances slower in early career, an unsurprising finding given that it often takes time to become established as a scientist. A single paper early in a scientist's career could, for example, significantly reduce m. If one really wants to study rate of achievement, career progression of h or h_T should ideally be represented graphically, as in Figure 3, rather than using simple statistics such as parameter m.

Being similar in its characteristics to h, the same caveats apply when using the h_T index to assess publication performance. Distortion due to self-citation is the most obvious example [SCHREIBER, 2007; VINKLER, 2007]. The number of contributing authors is also an issue. Highly cited papers tend to have the most authors [KOSTOFF, 2007], suggesting relatively smaller individual contributions, particularly from persons well down the author list. A modified version of the h-index, h_I, was introduced by BATISTA & AL. [2006] that applied a weighting for the number of authors (A): $h_I = h^2/A$. Further work would be required to incorporate such a weighting for h_T. As mentioned above, citation rates differ between research fields. Indices such as h and h_T should not therefore be used to compare the research performance *per se* of individuals in contrasting disciplines.

As well as scoring citations from the most highly cited articles, the new h_T index also scores those papers that have too few citations to be included in the h-index count (i.e., it includes citations that fall below the Durfee square in a Ferrers diagram). VANCLAY [2007] suggested that one advantage of the h-index is that it truncates the tails of distributions, suggesting that errors in the construction of citation statistics tend to occur in the tails. Our view is that it is advantageous to score citations of the most highly cited articles. As for the tail associated with low-cited articles, we argue that it is again correct to score "sleeping beauties" and indeed all of the articles that an author has published. Of course, errors occur in compiling the necessary statistics; but the citations associated with publications with insufficient citations to contribute to the h-index are generally low. For the data relating to the six scientists that we investigated, for example, these citations contributed an average of only 17% of the total citation count.

Based on our findings, should the scientific community forgo the h-index in favour of h_T? We argue that h_T is a more appropriate means of scoring publication output because it both encompasses all citations and increases smoothly from year to year, as opposed to the somewhat irregular changes seen in h. Nevertheless, h has its advantages, most notably ease of determination (although note that we computed both indices in a simple Microsoft Excel spreadsheet, available on request from the first author) and conceptualisation. An h-index of 15 immediately conveys that an individual has 15 papers, each with at least 15 cites. Yet if instead one proudly announces an h_T score of 30, it is much less clear what this means.

At least qualitatively, h and h_T are remarkably alike. If the two indices had shown different trends, as seen in the career progression of the scientists that we studied, then we would have no hesitation in recommending h_T as the preferred choice for measuring the publication output of scientists. Given the empirical nature of the analysis, and the close correlation between h and h_T, one could argue that either may be successfully employed, both being valid and useful bibliometric measures of performance.

The reader is therefore left with the choice of using h_T, which is theoretically superior and shows smooth progression, or the conceptually simple h as originally proposed by HIRSCH [2005].

*

TRA, RKSH and PDK are funded by the Natural Environment Research Council, U.K. We wish to thank Prof. A.P. Roberts for his comments on the manuscript, and Prof. J.G. Shepherd for suggesting the name "tapered h-index" for our new metric.

References

ANDREWS, G. E. (1984), *The Theory of Partitions*. Cambridge University Press.

BALL, P. (2005), Achievement index climbs the ranks, *Nature*, 448 : 737.

BATISTA, P. D., CAMPITELI, M. G., KINOUCHI, O., MARTINEZ, A. S. (2006), Is it possible to compare researchers with different scientific interests? *Scientometrics*, 68 : 179–189.

BORNMANN, L., DANIEL, H.-D. (2005), Does the *h*-index for ranking of scientists really work? *Scientometrics*, 65 : 391–392.

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 : 1275–1278.

EGGHE, L. (2006), Theory and practise of the *g*-index, *Scientometrics*, 69 : 131–152.

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 : 16569–16572.

HOLDEN, C. (2005), Random samples: Data point – Impact factor, *Science*, 309 : 1181.

JIN, B. H. (2006), H-index: an evaluation indicator proposed by scientist, *Science Focus*, 1 : 8–9.

KOSTOFF, R. N. (2007), The difference between highly and poorly cited medical articles in the journal Lancet. *Scientometrics*, 72 : 513–520.

LIANG, L. (2006), H-index sequence and h-index matrix: Constructions and applications, *Scientometrics*, 69 : 153–159.

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*, 58 : 297–301.

OPTHOF, T. (1997), Sense and nonsense about the impact factor, *Cardiovascular Research*, 33 : 1–7.

OPTHOF, T., CORONEL, R., PIPER, H. M. (2004), Impact factors: no *totum pro parte* by skewness of citation, *Cardiovascular Research*, 61 : 201–203.

REDNER, S. (2005), Citation statistics from 110 years of Physical Review, *Physics Today*, 58 : 49–54.

SAHA, S., SAINT, S., CHRISTAKIS, D. A. (2003), Impact factor: a valid measure of journal quality? *Journal of the Medical Library Association*, 91 : 42–46.

SCHREIBER, M. (2007), Self-citation corrections for the Hirsch index, *Europhysics Letters*, 78 : 1–6.

SEGLEN, P. O. (1992), The skewness of science, *Journal of the American Society for Information Science and Technology*, 43 : 628–638.

VAN RAAN, A. J. F. (2004), Sleeping beauties in science, *Scientometrics*, 59 : 467–472.

VANCLAY, J. K. (2007), On the robustness of the *h*-index, *Journal of the American Society for Information Science and Technology*, 58 : 1547–1550.

VINKLER, P. (2007), Eminence of scientists in the light of the *h*-index and other scientometric indicators, *Journal of Information Science*, 33 : 481–491.