J Math Chem (2010) 47:548–550
DOI 10.1007/s10910-009-9573-x

LETTER TO THE EDITOR

A monodimensional scientific performance measure: the *h* index, can be substituted by simple multidimensional descriptors?

Ramon Carbó-Dorca

Received: 8 January 2009 / Accepted: 1 July 2009 / Published online: 25 July 2009

Sir:

It is obvious that the lifelong work of a scientist, even if young, cannot be just described by a number. To fully accept this reductionist way is in my opinion outrageous. Yet, nowadays it appears that some descriptions of scientific performance (SP) only rely on such a scalar measure mentality.

In order to have a typical sample of this possible way of measuring SP, the interested or concerned reader just has to peruse the following website [http://es.geocities.com/indice_h/7], where the *h* index [1] of top *h*-indexed Spanish scientists is depicted. This website constitutes a local illustration to grasp how *h* index can vary, but appears to be sufficiently adequate and encompassing a large number of scientific disciplines for the purpose to have a reliable example. Other studies have obviously appeared; see for example Ref. [2], where high performance *h* scores, outnumbering sometimes in about two orders of magnitude the modest rankings of reference [http://es.geocities.com/indice_h/7], for various international top scientists is depicted.

Scalar measures are the usual means of ranking human performances as, for instance, in the sport field. However, individual's science performance do not looks even similar to a sport mark, but has at least the add-on purpose of increase knowledge about the world as exempt of error as possible, among other differential characteristics, when compared with other kinds of simpler human achievement.

Anybody with some experience will agree on that it is necessarily needed something slightly more complicated than a simple scalar in order to make a picture of any branch dedicated to increase knowledge. But sorrowfully enough, when searching for SP in the usual websites, one can feel that just a simple numerical artefact is needed to scale

R. Carbó-Dorca (✉)
Institut de Química Computacional, Universitat de Girona, Campus de Montilivi,
17071 Catalonia, Girona, Spain
e-mail: quantumqsar@hotmail.com

the human persons who apply for some administration money, intra or inter university promotion or any kind of similar stuff in appropriate places. Perhaps such a simple minded panorama is artificially created by the political pressure, originating within the usual bureaucratic scenarios worldwide. Moreover, more serious studies about SP have appeared recently, see Ref. [3], where the SP citation schemes are criticized in deep and studied with surprising results.

Most surely, the best way to obtain a better point of view about scientific work can adopt the idea that scalar measures shall be transformed into some sort of a multidimensional formalism, that is: into a set of several ordered scalars, even if such an objective requires that administrative desks should have access to an adequate program, which can transform or project multidimensional data into graphical or numerical form, in order to interpret the final SP output for any useful purpose. In fact, in the present times there are such devices on line, which can be more or less easily consulted by anyone and that could be straightforwardly reprogrammed to offer new and more sophisticated insight on collective or individual scientific professional careers. One of such possibilities has been successfully exploited in Ref. [3]. The references which can be followed in the previous reference lead to interesting websites, like the one in Ref. [4] where a quite serious and far reaching point of view about SP assessment is done.

In order to start the SP assessment with alternative simple new ways, first one can analyze the structure of the h index, a well documented SP proposal [1], which measures it in a plausible way. But, for instance, h index do not tells anything about the citation magnitude of the starting paper of the sequence leading to h itself. This is surprising enough, as one can imagine, among other possibilities, that there can appear SP measure sub-products, where even higher h rates correspond to some poor citation accumulation.

As a way to easily discuss the intention of the present proposal, the following notation will be used. Suppose that $K = \{K_I\} \subset \mathbb{N}$ is the citation sequence (CS) of a given author, ordered in such a way that: $K_0 \tilde{>} K_1 \tilde{>} \cdots K_N$. The index h is defined as: $h = K_h$ or something alike, as there is not any need to argue about it. To illustrate the h index performance, suppose some ideal case, where a researcher has the CS K made as: $\forall I$: $K_I = N - I$, which implies $h = N/2$. Suppose another CS, where $K_0 = pN$ with $p > 1$, but after this, the CS appears to be as before, then the same h index appears to be given to both. An effortless way to arrive to a two dimensional (2D) SP description could be obtained by means of the construction of the ordered pair: $\langle B| = (K_0\, h)$. Such a Cartesian product will describe individual SP much better than solely one of the two implied elements. In this manner, a set of scientists could be described as a positive quadrant set of points. Then, appearance of clusters, isolated points, or other substructures can be visually manifested in such pictograms. Although in a more elaborated way, a similar idea has been employed in a systematic study of a large amount of data concerning Physics in Ref. [3].

Moreover, 2D Minkowski, Euclidean or other kinds of distances between pairs of scientists can be easily computed, acting as dissimilarity measures between them. Appropriate similarity measures can be easily set; for instance, using the squared cosine between two ordered pairs or simply an appropriate inverse transformation of distances. Thus, ordering of a set of $\langle B|$ couples can be effortlessly obtained from dissimilarity or similarity measures, as these previously described. Such an ordering

can create a superposed tree or dendogram on any researcher set described in such a 2D way. Thus, a scientist can possess a relative order in the set of scientists of the same knowledge branch. Another refined albeit simple SP measure could be set by computing the sum: $\sigma_h = \sum_{I=0}^{h} K_I$; the ordered pair: $(\sigma_h\, h)$ can act as another 2D SP index.

The same or almost the same remarks as given before can be adopted in this case. More elaborate structures can be set up, just employing N-dimensional sets of parameters. For example, one can take into account that any CS K can originate a probability distribution, being a set of absolute frequencies, a non-empty natural number sequence. One can write in any circumstance: $T = \sum_I K_I \wedge P_I = K_I/T \rightarrow \sum_I P_I = 1$. The citation probability distribution (CPD): $P = \{P_I\} \subset \mathbb{Q}$ can be also taken as a first step for a well defined set of SP measures, both for individual and comparative purposes. Usual statistical techniques can be employed for comparison tasks: a well known collection of algorithms are known from old, see for example [5]. Furthermore, the h index can be associated to a percent of citation up to the h cutting edge, using the effortlessly computed sum: $\pi_h = \sum_{I=0}^{h} P_I$, so a new couple of ordered numbers can be employed: $(\pi_h\, h)$, to assess SP in a 2D scenario, as these already described. Entropy can be easily constructed from the CPD. Surely may be by itself a good SP index, which can be added to the previously 2D defined ones, structuring perhaps a three dimensional SP descriptor.

Despite the apparently deep analysis which has lead to the adoption of h index as a SP measure, surely that there are many ways 2D or higher dimensional to describe refined SP indices. They will be much more reliable and sound, without losing simplicity but adding sharpness into the task to assess SP.

From the scientific and administrative points of view, these alternatives or other equally sound or even better, sooner or later have to be seriously and globally considered in order to obtaining a coherent and reliable manner to construct SP.

Acknowledgements The altruistic kindness of Professor D. Gernert has permitted to enlarge this letter in a significant way. The partial sponsoring of the Ministerio of Educación y Ciencia by means of the fellowship: CTQ2006-04410/BQU is highly acknowledged.

References

1. J.E. Hirsch, Proc. Nat. Acad. Sci. **102**, 16569 (2005)
2. Ph. Ball, Nature **448**, 737 (2007)
3. S. Maslov, S. Redner, J. Neurosci. **28**, 11103 (2008)
4. R. Adler, J. Ewing, P. Taylor, *Citation Statistics*. (Report from the International Mathematical Union (IMU), 2008), http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf
5. L. Sachs, *Applied Statistics* (Springer, New York, 1982)