Criticism on the hg-index

FIORENZO FRANCESCHINI [1] AND DOMENICO MAISANO[1]

POLITECNICO di TORINO

([1]) Dipartimento di Sistemi di Produzione ed Economia dell'Azienda (DISPEA)

Corso Duca degli Abruzzi 24, 10129 - Torino, ITALY,

Tel. +39 011 5647225, Fax. +39 011 5647299, e-mail: fiorenzo.franceschini@polito.it

Abstract

Although composition of bibliometric indicators appears to be desirable, in many cases it may be misleading. After a brief introduction on the properties of scales of measurement, the attention of this communication is focused on a recent composite indicator, the *hg*-index, suggested by Alonso et al. (2010). Specifically, *hg*-index has three major criticalities: (1) the *hg* scale is the result of a composition of the *h*- and *g*-indices, which are defined both on ordinal scales, (2) the equivalence classes of *hg* are questionable and the substitution rate between *h* and *g* may arbitrarily change depending on the specific *h* and *g* values, (3) the apparent increase in granularity of *hg*, with respect to *h* and *g*, is illusory and misleading. Argument is supported by several examples.

Keywords: bibliometrics, *hg*-index, *h*-index, *g*-index, indicator composition, composite indicator, scales of measurement, ordinal scale, scale granularity.

1. Introduction

Many different bibliometric indicators can be used to evaluate the scientific production of a scientist or a scientific journal. It is often recommended using a set of indicators, so as to take many aspects into account and provide an exhaustive picture of the matter of interest [Costas and Bordons, 2007, Franceschini et al., 2007]. Undoubtedly, synthesis or aggregation of several indicators is an operation that may simplify the analysis. For example, a relatively recent but very popular synthetic indicator is the *h*-index, which synthetically aggregates two important aspects of the output of a scientist: diffusion/impact – represented by the number of citations per paper – and productivity – represented by the number of different papers [Hirsch, 2005]. Determining *h* is very simple: papers have to be sorted in decreasing order according to the citations they received and then to be counted, stopping the tally at the breakeven point between the citation number and the number of examined papers (see the first two columns of the table in Fig. 1). The synthesis capability, together with the immediate intuitive meaning and robustness are the basic reasons of the great diffusion of *h* over the scientific community. This success is also confirmed by the appearance of a huge number of proposals for new variants and improvements of *h* (such as *g*, *AR*, *m*, h_w, and many others). For more on the merits and weak points of *h* and the large number of proposals for new variants and improvements, we refer the reader to the vast literature and extensive reviews [Braun et al., 2006; Glänzel, 2006; Rousseau, 2008; Egghe, 2010; Franceschini and Maisano, 2010a; Franceschini and Maisano, 2010b].

citations for each paper	rank	cumulative no. of citations	rank^2
30	1	30	1
20	2	50	4
18	3	68	9
12	4	80	16
9	5	89	25
8	6	97	36
8	**7**	105	49
6	8	111	64
6	9	117	81
5	10	122	100
4	**11**	126	121
3	12	129	144
2	13	131	169
2	14	133	196
2	15	135	225
…	16	…	256

h-core

g-core

Fig. 1 – Example of calculation of the *h*-index (defined as *the number such that, for a general group of papers, h papers received at least h citations while the other papers received no more than h citations* [Hirsch, 2005]) and *g*-index (defined as *the maximum number of a scientist's most cited papers, so that they have together at least g^2 citations* [Egghe, 2006]). In this specific case *h*=7 and *g*=11.

We define as indicator composition the synthesis of two or more other indicators by arithmetic or mathematical operations (i.e. addition, multiplication, power, etc…). In general, when composing indicators, their measurement scales must be taken into account very carefully [Roberts, 1979].

The goal of this communication is to underline the risks of the mathematical composition of bibliometric indicators, being aware of the fact that, very often, it represents a dangerous temptation. In particular, the attention is focused on a recent indicator, the *hg*-index, introduced by Alonso et al. (2010), and defined as the geometric mean of the *h*- and *g*-indices of a researcher:

$$hg = \sqrt{h \cdot g} \qquad (1)$$

The ambitious objective of *hg*-index is trying to fuse all the benefits and minimise the disadvantages of *h* and *g* [Alonso et al., 2010]. Precisely, a significant disadvantage ascribed to *h* is that it is insensitive to the so called “big hit” papers, that is to say those papers with outstandingly high number of citations [Hirsch, 2005; Egghe, 2006; Jin et al., 2007]. On the other hand, the greatest drawback ascribed to *g* is that it may be greatly influenced by a very successful paper.

We point out that the largest part of the *h* variants are defined and constructed starting from the list of papers and citations of a scientist. Thus, they can be defined as 1st level indicators. *hg*, which is a composition of two 1st level indicators (i.e. *h* and *g*), can be seen as a 2nd level indicator (see Fig. 2).

In our opinion, *hg* has three basic criticalities that will be analysed in detail: (1) *hg* derives from a composition of two indicators, *h* and *g*, defined on distinct ordinal scales, (2) the equivalence classes of *hg* are questionable and the substitution rate between *h* and *g* may arbitrarily change depending on the specific *h* and *g* values, (3) the apparent increase in granularity with respect to *h* and *g* is illusory and misleading.

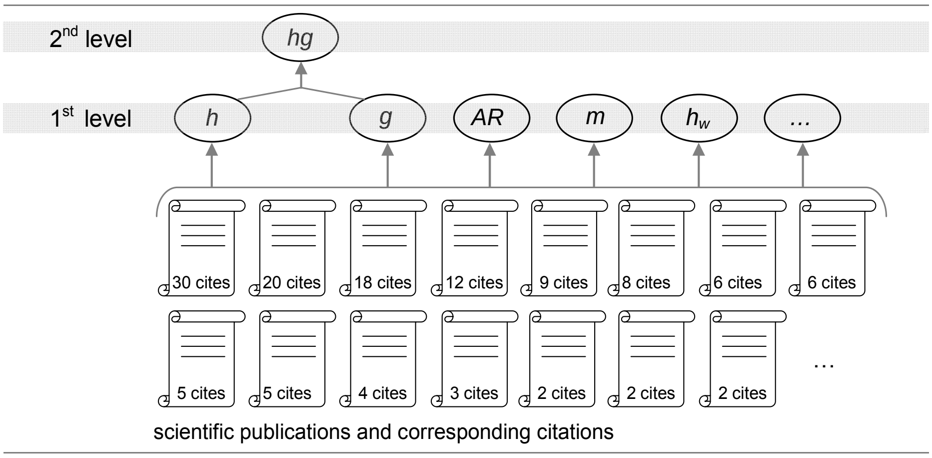

Fig. 2 – Interpretation of *hg* as a 2nd level indicator: *hg* aggregates two 1st level indicators (i.e. *h* and *g*), which are constructed starting from the list of papers and citations of a scientist.

The remaining of this communication is organised into four sections. The first one contains a brief overview of the properties of the scales of measurement, while the other three are associated to the three critical aspects of *hg* mentioned before. Argument is supported by several examples. Most of the considerations about *hg* can be extended to other composite indicators, which are based on the mathematical composition of other 1st level indicators – for instance the q^2-index, suggested by Cabrerizo et al. (2010).

2. Basic properties of the scales of measurement

A largely accepted classification of the scales of measurement was proposed by Stevens (1946). In this proposal, measurements/indicators can be classified into four different types of scales: nominal, ordinal, interval and ratio (see Tab. 1).

Scale Type	Empirical Properties	Permissible Statistics	Examples
Nominal	Equivalence	Mode, chi square	Eye colour, place of birth, etc…
Ordinal	Equivalence, order (greater or less)	Median, percentile	Surface hardness, military rank, etc…
Interval	Equality, order, distance (addition or subtraction)	Mean, standard deviation, correlation, regression, analysis of variance	Temperature in °C, serial numbers, etc…
Ratio	Equality, order, distance, ratio (multiplication or division)	All statistics permitted for interval scales plus the following: geometric mean, harmonic mean, coefficient of variation, logarithms	Temperature in K, weight, age, number of children, etc…

Tab. 1 – Classification scheme of measurements/indicators depending on their scale types [Stevens, 1946; Roberts, 1979].

It will be convenient to illustrate this scheme with a variable X and two objects, say A and B, whose scores on X are x_A and x_B, respectively.

1. A nominal scale merely distinguishes between classes. That is, with respect to A and B one can only say $x_A = x_B$ or $x_A \neq x_B$.

2. An ordinal scale induces an ordering of the objects. In addition to distinguishing between $x_A = x_B$ and $x_A \neq x_B$, the case of inequality is further refined to distinguish between $x_A > x_B$ and $x_A < x_B$.
3. an interval scale assigns a meaningful measure of the difference between two objects. One may say not only that $x_A > x_B$, but also that A is $x_A - x_B$ units different than B.
4. a ratio scale is an interval scale with a meaningful zero point. If $x_A > x_B$ then one may say that A is x_A / x_B times superior to B.

From the viewpoint of the scale properties, the above types of measurement scales are ordered from "less powerful" to "more powerful". In particular, the more powerful scales (interval and ratio) provide more information and are generally preferred for measurement purposes. It is often a goal of measurement to obtain scales that are as much powerful as possible, but – unfortunately – this is not always so straightforward.

As a general rule, numbers should be analysed on the basis of the properties of the scale with which they are gathered [Roberts, 1979]. Consequently, one may obtain results that do not make sense by applying arithmetic operations to measurements/indicators with scales in which these operations are inadmissible (see the second column of Tab. 1).

3. *h* and *g* measurement scales

h and *g*, although obtained by different approaches, both represent the number of a scientist's most cited papers (i.e. *h*- and *g*-core in Fig. 1). For these indicators, the link between the indicator value and one scientist's papers and citations is immediate, due to the simplicity of their definitions. An *h*-index of 10 immediately conveys that an individual has 10 papers, each with at least 10 cites, while a *g*-index of 10 conveys that the 10 most cited papers have, together, at least 10^2 citations [Hirsch, 2005; Egghe, 2006]. But what can we say about their measurement scales? To make the analysis clearer, let now focus the attention on *h*. However, the following considerations can be almost integrally extended to *g*.

Although being expressed by natural numbers, *h* is defined over an ordinal scale with only equivalence and ordering properties. So, only comparisons of ordering (greater and less) can be made, in addition to equivalence. For the purpose of example, if two scholars (A and B) have the same *h*, they are considered as equivalent, while, if $h_A > h_B$, then A is considered better than B. In such a scale, operations like conventional addition and subtraction are without meaning, since equal differences between *h* values do not necessarily represent equivalent intervals [Roberts, 1979]. That means that it is not possible to say how much A is better than B, according to the difference $h_A - h_B$, regardless of the values of h_A and h_B [Franceschini and Maisano, 2010]. An indirect demonstration is that, for high values of the *h*-index, it becomes more and more difficult to increase it [Egghe, 2007; Burrell, 2007]. In other words, the "gap" between two scientists with *h*-indices 5 and 7 is much larger than the gap between two scientists with h-indices 2 and 4. More precisely, considering a scholar's scientific production, Hirsch empirically showed that, on the average, the total number of citations (C) is approximately proportional to h^2 [Hirsch, 2005]:

$$C \approx a \cdot h^2 \tag{2}$$

Thus, h value is roughly proportional to $C^{1/2}$. Considering a particular class (h), the distance from the higher consecutive class (h+1) – in terms of citations – can be calculated as:

$$C(h+1) - C(h) \approx a \cdot (h+1)^2 - a \cdot h^2 = a \cdot (2h+1) \tag{3}$$

So, Eq. 3 shows that the average distance between two consecutive (h) classes increases proportionally with the h value. For example, if $h = 5$ this distance is about $a \cdot 11$, and if $h = 10$ it is about $a \cdot 21$ citations.

All the previous remarks can be extended, with little modifications, to g. Consistently with the theoretical model proposed by Egghe (2006), an empirical relationship is present between C and g^2:

$$C \approx b \cdot g^2 \tag{4}$$

The difference between (Eq. 2) is that, in general, the proportionality coefficient between C and h^2 (coefficient a) is significantly higher than that one between C and g^2 (coefficient b). For example, we have determined a and b through a linear regression for a sample of more than 400 scientists, randomly selected in the field of Quality Engineering/Management. Results are $a \approx 3.85$ ($R^2 \approx 0.92$) and $b \approx 1.67$ ($R^2 \approx 0.86$). Similar results can be obtained considering scientists of other disciplines [Batista et al., 2006].

Considering one citation as the elementary unit of effort, it can be said that the distance between two consecutive graduations of the h and g scales tend to increase with their values.

To avoid any misunderstanding, we remark the fact that Eqs. 2, 3 and 4 tend to be respected *on the average*. Thus, in terms of average citations, the h and g scales can be represented as illustrated in Fig. 3 (g-h plan).

Being defined as the geometric mean of two natural numbers, which are conveyed on ordinal scales, the hg's scale is meaningless because it is determined by an operation that is not formally permitted. As seen in Tab. 1, ordinal scales do not admit arithmetic operations of addition, subtractions, multiplication or division and such relationships are inadmissible when dealing with ordinal data [Stevens, 1948; Roberts, 1979]. If we wanted to be more precise, it could be said that the h and g scales not only have the ordering property, but – on average – they are referable to interval scales with non-equidistant graduations. Despite this, the composition performed by hg is still inadmissible.

Finally, as the Alonso et al. (2010) admit themselves, another remarkable fact is that hg is a number with no direct meaning in terms of papers and citations of a scientist.

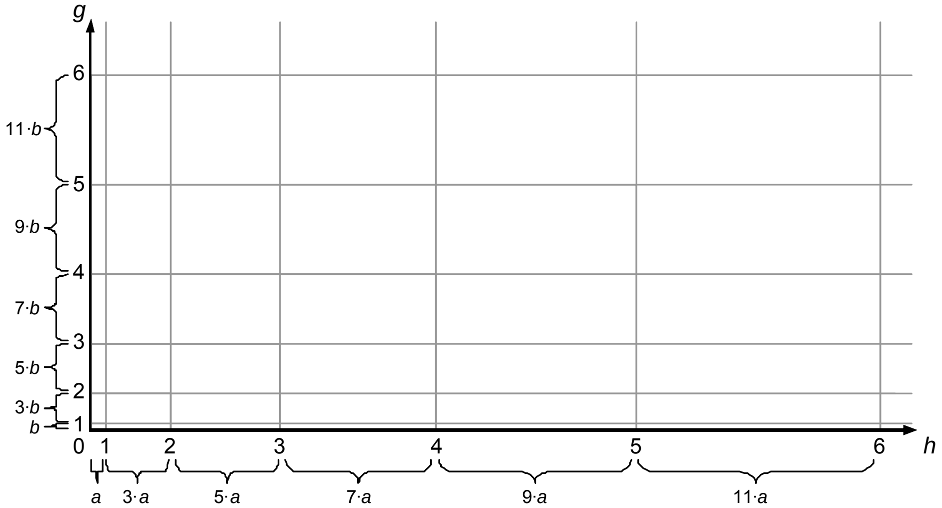

Fig. 3 – *g-h* plan. The distance – in terms of citations – between two consecutive *h* and *g* graduations tends to increase with the corresponding *h* and *g* values, respectively with the laws *a*·(2*h*+1) and *b*·(2*g*+1) (being *a*>*b*). To emphasize this fact, *h* and *g* axes are deformed accordingly. This representation makes sense only under the assumption of average proportionality among h^2, g^2 and *C*.

4. *h* and *g* equivalence classes and substitution rate

Fig. 4 is an example of a *g*-*h* plan, constructed considering the same sample of more than 400 scientists mentioned before. As a direct consequence of Eq. 2 and Eq. 4, points tend to be distributed along the line of equation:

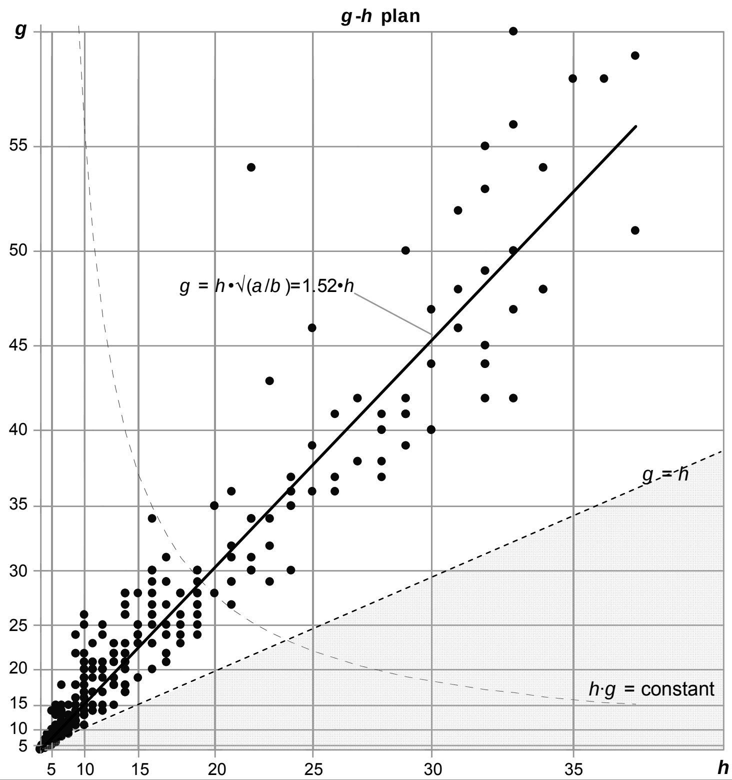

Fig. 4 – Empirical relationship between *g* and *h*. Graph is constructed considering a sample of 422 scientists, randomly selected in the area of Quality Engineering/Management. The distance – in terms of average citations – between two consecutive *h* and *g* graduations tends to increase with the corresponding *h* and *g* values, respectively with the laws *a*·(2*h*+1) and *b*·(2*g*+1) (being *a*>*b*). To emphasize this fact, *h* and *g* axes are deformed accordingly. It can be seen that points tend to be distributed along the line of Eq. 5. Since *g* ≥ *h*, the area below the bisector (*g*=*h*) is empty. Also, it can be noticed that the bisector does not form a 45 degree angle to the horizontal line.

$$g = \sqrt{\frac{a}{b}} \cdot h \tag{5}$$

The portion of the *g*-*h* plan below the bisector (highlighted in grey) is empty (being, in general, $g \geq h$) [Egghe, 2006; Alonso et al., 2010].

Iso-*hg* curves, that is the geometric loci of the points with hg-index = constant ($h{\cdot}g$ = constant), appear as some hyperbolae. Now, the question is: what is the rationale for considering two points laying on the same hyperbola as equivalent? Let analyse the problem through an example (see Fig. 5).

Scientist A

$h_A=2$; $g_A=8$

citations for each paper	rank	cumulative no. of citations	rank^2
29	1	29	1
29	2	58	4
1	3	59	9
1	4	60	16
1	5	61	25
1	6	62	36
1	7	63	49
1	8	64	64
1	9	65	81
1	10	66	100
1	11	67	121
1	12	68	144
1	13	69	169
1	14	70	196
1	15	71	225
1	16	72	256
1	17	73	289

$hg_A = \sqrt{2 \cdot 8} = 4$

$h_A > h_B$
$g_A < g_B$
$hg_A = hg_B$

Scientist B

$h_B=1$; $g_B=16$

citations for each paper	rank	cumulative no. of citations	rank^2
241	1	241	1
1	2	242	4
1	3	243	9
1	4	244	16
1	5	245	25
1	6	246	36
1	7	247	49
1	8	248	64
1	9	249	81
1	10	250	100
1	11	251	121
1	12	252	144
1	13	253	169
1	14	254	196
1	15	255	225
1	16	256	256
1	17	257	289

$hg_B = \sqrt{1 \cdot 16} = 4$

Fig. 5 – Comparison of two (fictitious) scholars with the same *hg*-index, but different *h* and *g* values.

Scientist A has 17 papers: two with 29 citations each and the remaining with one citation only. Scientist B has 17 papers as well: one with 241 citations and the remaining with one citation only. *hg*-indices are coincident, even if $C_B = 257$ is more than three times larger than $C_A = 73$. The synthesis of *h* and *g* by means of *hg* introduces a questionable equivalence (i.e. $hg_A = hg_B = 4$) that subverts both the classifications according to *h* (i.e. h_A=2 > h_B=1) and *g* (i.e. g_A=8 < g_B=16).

Also, it can be noticed that the substitution rate – defined as the rate at which the *h*-index can be increased/decreased in exchange for a decrease/increase in the *g*-index, maintaining the same *hg* value – is not constant. Assuming that *h* and *g* are defined on ratio scales, the substitution rate would be:

$$\left(h \cdot g = \text{constant}\right) \rightarrow \frac{\Delta g}{\Delta h} = -\frac{\text{constant}}{h^2} \qquad 6)$$

which can be geometrically represented by the angular coefficient of the line tangent to the iso-*hg* curve. Eq. 6 says that this quantity is not constant over the *h* domain, since it depends on the scientist's *h* and *g* position. What is the rationale behind?

5. Granularity does not necessarily mean precision

Another presumed advantage of *hg* (defined over the domain of real positive numbers $\mathbb{R}_0^+$) would be the higher granularity with respect to *h* and *g* (both defined over the domain of natural numbers $\mathbb{N}_0$) [Alonso et al., 2010]. Despite this, when scientists with very close *h* and *g* values are compared and it is not easy to determine an overall classification, using *hg* to sort things out can be dangerous. In fact, the higher granularity of the *hg*'s measurement scale is just the result of an improper alteration in the *h* and *g* scales, but it does not necessarily mean higher discrimination power. By comparison, it would be like calculating the arithmetic mean of two measurements of the same mass, taken using two low resolution measuring instruments (for example two balances with resolution of 5 grams) and keeping many significant figures in the resulting number (for example, assuming the result expressed in grams with two digits after the decimal point), overstating the resolution of the measurement. In other words, we think that the illusion of having created a new indicator with a higher discriminatory power, synthesising the information provided by two indicators (*h* and *g*) with lower granularity, can increase the risk of passing hasty judgements. Sometimes, we must accept the fact that some situations cannot be distinguished on the basis of the observation tool(s) on hand [Franceschini et al., 2007].

6. Concluding remarks

This communication analyses some drawbacks of the *hg*-index, which is obtained by the geometric mean of *h* and *g*. In particular, the discussion is focused on the risks that this composition may introduce.

In general, when dealing with bibliometric indicators with distinct scales (like *h* and *g*) we think that the most transparent way to use them together is to draw a map (see the *g*-*h* plan in Fig. 4) illustrating the bibliometric positioning of different scientists. Even if a map is unable to give a unique synthesis – this will often be a "forbidden dream" – it can be useful and does not cause any alteration of the indicators of interest.

References

Alonso, S., Cabrerizo, F., Herrera-Viedma, E., Herrera, F. (2010) *hg-index: a new index to characterize the scientific output of researchers based on the h- and g-indices*, Scientometrics, vol. 82, n. 2, pp. 391-400, DOI: 10.1007/s11192-009-0047-5

Batista, P.D., Campiteli, M.G., Kinouchi, O., Martinez, A.S. (2006) *Is it possible to compare researchers with different scientific interests?*, Scientometrics, vol. 68, n. 1, pp. 179-189

Braun, T., Glänzel, W., Schubert, A. (2006) *A Hirsch-type index for journals*, The Scientist, vol. 69 n. 1, pp. 169-173

Burrell, Q. L. (2007) *Hirsch index or Hirsch rate? Some thoughts arising from Liang's data*, Scientometrics, vol. 73, n. 1, pp. 19-28.

Cabrerizo, F.J., Alonso, S., Herrera-Viedma, E., Herrera F. (2010) *q^2-Index: Quantitative and qualitative evaluation based on the number and impact of papers in the Hirsch core*, Journal of Informetrics, vol. 4, n. 1, pp. 23-28

Costas, R., Bordons, M. (2007) *The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level*, Journal of Informetrics, vol. 1, n.3, pp. 193-203

Egghe, L. (2006) *Theory and practise of the g-index*, Scientometrics, vol. 69, n. 1, pp. 131-152

Egghe, L. (2007) *Dynamic h-index: the Hirsch index in function of time*, Journal of the American Society for Information Science and Technology, vol. 58, n. 3, pp. 452-454

Egghe, L. (2010) *The Hirsch-index and related impact measures*, Annual Review of Information Science and Technology, vol. 44, p. 65-114

Franceschini, F., Galetto, M., Maisano, D. (2007) *Management by Measurement: Designing Key Indicators and Performance Measurement Systems*, Springer Verlag, Berlin

Franceschini, F., Maisano, D. (2010a) *Analysis of the Hirsch index's operational properties*, European Journal of Operational Research, vol. 203, n. 2, pp. 494-504, DOI: 10.1016/j.ejor.2009.08.001

Franceschini, F., Maisano, D. (2010b) *Analysis of the ch-index, an indicator to evaluate the diffusion of scientific research output by citers*, to appear in Scientometrics, DOI: 10.1007/s11192-010-0165-0

Glänzel, W. (2006) *On the opportunities and limitations of the h-index*, Science Focus, vol. 1, n. 1, pp. 10-11

Hirsch, J.E. (2005) *An index to quantify an individual's scientific research output*, in Proceedings of the National Academy of Sciences of the United States of America, vol. 102, pp. 16569-16572

Jin, B.H., Liang, L.M., Rousseau, R., Egghe, L. (2007) *The R- and AR-indices: Complementing the h-index*, Chinese Science Bulletin, vol. 52, n. 6, pp. 855-963

Roberts, F.S. (1979) *Measurement theory: with applications to decisionmaking, utility, and the social sciences*, Encyclopedia of Mathematics and its Applications, vol. 7, Addison-Wesley, Reading, MA

Rousseau, R. (2008) *Reflections on recent developments of the h-index and h-type indices*, COLLNET Journal of Scientometrics and Information Management, vol. 2, n. 1, pp. 1-8

Stevens, S.S. (1946) *On the Theory of Scales of Measurement*, Science, vol. 103, n. 2684, pp. 677-680, DOI: 10.1126/science.103.2684.677