

Characteristic scores and scales based on h-type indices

Peer-reviewed author version

EGGHE, Leo (2010) Characteristic scores and scales based on h-type indices. In:
JOURNAL OF INFORMETRICS, 4(1). p. 14-22.

DOI: 10.1016/j.joi.2009.06.001

Handle: <http://hdl.handle.net/1942/9684>

Characteristic scores and scales based on h -type indices

by

L. Egghe

Universiteit Hasselt (UHasselt), Campus Diepenbeek, Agoralaan, B-3590 Diepenbeek,

Belgium

leo.egghe@uhasselt.be

ABSTRACT

Based on the rank-order citation distribution of e.g. a researcher, one can define certain points on this distribution, hereby summarizing the citation performance of this researcher. Previous work of Glänzel and Schubert defined these so-called “Characteristic Scores and Scales” (CSS), based on average citation data of samples of this ranked publication-citation list.

In this paper we will define another version of CSS, based on diverse h -type indices such as the h -index, the g -index, the Kosmulski’s $h^{(2)}$ -index and the g -variant of it, the $g^{(2)}$ -index.

Mathematical properties of these new CSS are proved in a Lotkaian framework. These CSS also provide an improvement of the single h -type indices in the sense that they give h -type index values for different parts of the ranked publication-citation list.

Keywords and phrases: characteristic scores and scales, CSS, h -index, g -index, R -index, $h^{(2)}$ -index, $g^{(2)}$ -index.

I. Introduction

The visibility or impact of a researcher's scientific output is measured using indicators based on the rank-order citation distribution of this researcher – stated more simply – on the ranked list of papers of this researcher, ranked in decreasing order of number of citations to each of these papers. A typical – and famous – example of such an indicator is the Hirsch-index or h -index (Hirsch (2005)). In subsequent years, several other, so-called h -type indices, were introduced in order to avoid certain “bad properties” such as insensibilities of the h -index. We can mention here the g -index (Egghe (2006a,b,c)), the R -index (Jin, Liang, Rousseau and Egghe (2007)), Kosmulski's $h^{(2)}$ -index (Kosmulski (2006)) and its g -index variant $g^{(2)}$ (Egghe (2009a)). We refer to Egghe (2009b) for an extensive review chapter on h -type indices.

Yet, all these indices suffer from the same aspect of roughness: they are all single indicators, hereby reducing the evaluation of a researcher's scientific career, to a single (i.e. one-dimensional) observation. One way to overcome this is to measure these indices year-per-year, hereby yielding a sequence of time-dependent h -type indices (see Egghe (2009b) – last section before the conclusions section and the references therein). This is an interesting approach yielding time-series of yearly impact, hence describing well the evolution of the impact of a researcher over the years (with a possibility to predict future impact of this researcher).

In the above approach we are still having the problem of the fact that, per year, we only have one impact indicator value. It would be interesting to describe, at any moment of our choice, the rank-order citation distribution of a researcher in a better way than just presenting one indicator. This problem, surprisingly in a certain sense, has been addressed more than 20 years ago in Glänzel and Schubert (1988). The general idea is to determine, on the rank-order citation distribution, a sequence of points ν_k , $k=1,2,\dots$ and β_k , $k=1,2,\dots$, where the ν_k are some ranks of papers and $\beta_k = \gamma(\nu_k)$ are the corresponding “characteristic scores”, i.e. the number of citations to the paper on rank $r = \nu_k$.

Here γ is the rank-order frequency function, where $\gamma(r)$ expresses the number of citations to the paper on rank r . The methodology is called the method of “Characteristic Scores and Scales” (CSS).

In the above definition the ν_k and β_k are arbitrary, but one interesting example is given in Glänzel and Schubert (1988) – see also Glänzel (2008). It boils down to the following, based on the ranked list of papers of a researcher, in decreasing order of number of citations received. First we take the average number of citations per paper, denoted β_1 . Now we discard all papers with less citations than this mean β_1 . This yields a rank ν_1 of the last paper (i.e. highest rank) that is not discarded. We denote by β_2 the average number of received citations by the papers on ranks $1, \dots, \nu_1$ (the non-discarded papers). Then we discard all papers with less than β_2 citations, yielding a rank ν_2 of the last non-discarded paper, and so on. We hence obtain an increasing sequence β_k and a decreasing sequence ν_k (finite in practice but possibly infinite in a theoretical setting with rank-order distribution functions).

In Egghe (2009c) we give a mathematical formula for these β_k (and ν_k) in a Lotkaian framework where we show that, simply, $\beta_k = \mu^k$ for all k where μ is the average number of citations per paper in the researcher’s rank-order list. We also present a logical variant of the above where “average” is replaced by “median”, where we also obtain an increasing sequence β_k and a decreasing sequence ν_k .

It is clear that, in determining CSS, we are not limited to averages or medians and even not to increasing β_k s and (corresponding) decreasing ν_k s, although we expect that any algorithm will yield monotonic β_k s and (of course reversely) monotonic ν_k s.

In this paper, characteristic scores and scales are defined using h -type indices such as the h -index itself and its variants: the g -index, $h^{(2)}$ -index and $g^{(2)}$ -index (definitions are given in the next section). The methodology is based on the fact that, certainly for high values of e.g. the h -index, many papers of this researcher (i.e. certainly all papers with $h-1$ or less citations) are discarded and this can be papers with many citations. In order to give these (highly cited) papers also a role in the impact assessment of this researcher, in the next

section, we present a method of calculating the h -index of the truncated set of papers, where we have discarded the first h papers (and the same technique for the other indices).

This gives us a sequence of h -type indices β_k and corresponding ranks ν_k , where the monotonicity of these sequences is reverse to the ones discussed above (based on averages or medians). Models for these CSS are given in a Lotkaian framework. We also calculate CSS for this author for the h -, g -, R -, $h^{(2)}$ - and $g^{(2)}$ -indices and give interpretations of the obtained data.

The paper ends with conclusions and suggestions for further research.

II. Definition and properties of CSS based on h-type indices

In the Appendix one finds Table 1 which is the list of papers of this author in decreasing order of their number of received citations.

This list was created on February 4, 2009 based on the Web of Science (WoS) data (including the (Social) Science Citation Index and Conference Proceedings). The first CSS that we will define is based on the h -index itself.

II.1 CSS based on the h-index

The h -index is the highest rank $r = h$ such that all papers on ranks $1, \dots, h$ have h or more citations (Hirsch (2005)). It is readily seen from Table 1 that $h = 17$ for this author on that date. So when we only use $h = 17$ as an indicator of impact, we disregard all other papers including papers that received 17, 16, 15, citations. Let it be clear: we do not want to include papers with few citations: it is a good robustness property of all h -type indices to disregard these papers but papers with lots of citations not belonging to the so-called h -core (the first h papers) should be used in the measurement of impact. This is even more clear in the case of prolific authors as e.g. given in Ball (2005) (see also Egghe (2009b)) where h -indices above

100 can occur. A single number $h = 100$ means that certainly all papers with 99 or less citations are disregarded which is very strange in the assessment of impact!

The CSS based on the h -index is constructed as follows (explained on the data in the Appendix). We have $h_0 = h = 17$ in Table 1. Then we cut-off these 17 papers and start re-labeling the other papers from rank 1 onwards. We now calculate the h -index h_1 of this new list (see Table 2). We see that we have now $h_1 = 12$ (original rank 29). Then we cut-off these 12 papers and start relabeling the other papers from rank 1 onwards. The h -index h_2 for this new list can be calculated. In case of Table 1 we see that $h_2 = 9$, and so on. The values h_0 , $h_0 + h_1$, $h_0 + h_1 + h_2$, ... represent the ranks ν_k , $k = 1, 2, \dots$ described in the Introduction. Let us denote by $\gamma(r)$ the number of citations to the paper on rank r then we have that in this case $\gamma(\nu_k) = \beta_k$ are the characteristic scores.

It is also obvious that the sequence ν_k is increasing and the sequence β_k is decreasing, the opposite of the CSS in Glänzel and Schubert (1988), Glänzel (2008) and Egghe (2009c). The ranks ν_k delimit marks on the list of papers such that the papers on ranks $\nu_{k-1} + 1, \dots, \nu_k$ (define $\nu_0 = 1$) each have at least h_{k-1} citations but (if $k \geq 2$) no more than h_{k-2} citations. This yields a clear classification of the author's papers in citation classes and each of these classes contain a maximal number of papers with this property (by definition of the h -index). To provide a model for these CSS we briefly repeat the simple Lotkaian model: for $j \geq 1$:

$$\varphi(j) = \frac{C}{j^\alpha} \quad (1)$$

$C > 0$, $\alpha > 1$ is Lotka's law which is equivalent with Zipf's law

$$\gamma(r) = \frac{B}{r^\beta} \quad (2)$$

$r \in [0, T]$ (T = total number of sources, i.e. here: cited papers) and where

$$B = \left(\frac{C}{\alpha - 1} \right)^{\frac{1}{\alpha - 1}} \quad (3)$$

$$B = T^{\frac{1}{\alpha - 1}} \quad (4)$$

$$\beta = \frac{1}{\alpha - 1} \quad (5)$$

This result appears in Egghe (2005), Exercise II.2.2.6 p.134 and also in Egghe and Rousseau (2006) where, in the Appendix, a complete proof is given. In the latter article one also proves that

$$h_0 = h = T^{\frac{1}{\alpha}} \quad (6)$$

in this Lotkaian framework (the proof of (6) readily follows from (2), (4) and (5) and the h -index definition $\gamma(h) = h$).

It is not easy to give a model for the h_k -values. For $k = 0$ we have of course (6). For $k = 1$ we have the following proposition.

Proposition 1: Supposing the law of Lotka (1), we have that h_1 satisfies

$$h = h_1^{1-\frac{1}{\alpha}} (h_1 + h)^{\frac{1}{\alpha}} \quad (7)$$

In case $\alpha = 2$ we have

$$h_1 = h \frac{\sqrt{5}-1}{2} \quad (8)$$

which is the Golden Section (cf. Gellart et al. (1975)) of a line piece of length h .

Proof: Taking away the first h articles results in a ranked list of articles with rank-frequency function (by 2))

$$\gamma_1(r) = \frac{B}{(r+h)^\beta} \quad (9)$$

$r \in [0, T-h] = \left[0, T - T^{\frac{1}{\alpha}}\right]$. The h -index h_1 of this truncated list is, by definition $\gamma_1(h_1) = h_1$

or, by (9)

$$\frac{B}{(h_1 + h)^\beta} = h_1$$

hence

$$B = h_1 (h_1 + h)^\beta$$

Using (4), (5) and (6) proves (7). To the best of my knowledge, equation (7) cannot be solved except for $\alpha = 2$. We have then from (7) the equation (in the unknown h_1)

$$h_1^2 + h_1 h - h^2 = 0$$

yielding, since $h_1 > 0$, equation (8). □

A Golden Section of a line piece AB is the point X on AB such that

$$\frac{|AB|}{|AX|} = \frac{|AX|}{|XB|} \quad (10)$$

($|\cdot|$ denotes the length of the line piece). Hence (8) is the length of the largest line piece AX if $|AB| = h$. The value is, approximately

$$\frac{\sqrt{5}-1}{2} \approx 0.618034 \quad (11)$$

If we put $x = \frac{h_1}{h}$ then it is readily seen that (7) is equivalent with the equation

$$x^\alpha + x^{\alpha-1} = 1 \quad (12)$$

which can be solved numerically for all values α . Equation (7) and (10) clearly show that $h_1 < h$, an evident fact from the definition of h_1 .

The calculation of h_2 is even more difficult but we are still able to derive some useful results. We have the following proposition.

Proposition 2: Supposing the law of Lotka (1), we have that h_2 satisfies

$$h = h_2^{\frac{1}{\alpha}} (h_2 + h + h_1)^{\frac{1}{\alpha}} \quad (13)$$

In case $\alpha = 2$ we have

$$h_2 = 0.47726h \quad (14)$$

$$h_2 = 0.7722229h_1 \quad (15)$$

and (14) and (15) are conform with (8). From these results it follows that $\frac{h_2}{h_1} > \frac{h_1}{h_0}$ (where

$h = h_0$), using (8), (11) and (15).

Proof: By (2) and since now we cut-off after rank $h + h_1$, we have the following rank-frequency function for the remaining sequence of papers

$$\gamma_2(r) = \frac{B}{(r+h+h_1)^\beta} \quad (16)$$

where $r \in [0, T-h-h_1]$. The definition of the h -index h_2 for this truncated list is $\gamma_2(h_2) = h_2$, hence

$$\frac{B}{(h_2+h+h_1)^\beta} = h_2$$

This yields, using (4), (5) and (6), the equation (13). In case $\alpha = 2$, (13) gives, since $h_2 > 0$

$$h_2 = \frac{-(h+h_1) + \sqrt{(h+h_1)^2 + 4h^2}}{2} \quad (17)$$

Since this equation contains the h -indices h , h_1 as well as h_2 we will further calculate h_2 in function of h only and in function of h_1 only. This can be done using equation (8), yielding

(14) (replacing h_1 by (8)) and (15) (replacing h by (8), i.e. $h = \frac{2h_1}{\sqrt{5}-1}$). We can readily check

that (8), (14) and (15) indeed satisfy the relation $\frac{h_2}{h} = \frac{h_2}{h_1} \frac{h_1}{h}$ as it should. \square

A further generalization of the above method leads to the general formula (for $k = 1, 2, 3, \dots$)

$$h_k = \frac{-\sum_{j=0}^{k-1} h_j + \sqrt{\left(\sum_{j=1}^{k-1} h_j\right)^2 + 4h_0^2}}{2} \quad (18)$$

($h = h_0$ by definition).

We now turn our attention to CSS based on the $h^{(2)}$ -index.

II.2 CSS based on Kosmulski's $h^{(2)}$ -index

The $h^{(2)}$ -index of Kosmulski (Kosmulski (2006)) is the highest rank $r = h^{(2)}$ such that all papers on ranks $1, \dots, h^{(2)}$ have $(h^{(2)})^2$ or more citations. The usefulness of Kosmulski's $h^{(2)}$ -index has been criticized e.g. in Egghe (2009b) since $h^{(2)}$ usually is very low and in any case much lower than the h -index: in the case of Table 1, this author's $h^{(2)}$ -index equals $h^{(2)} = 6$, much smaller than $h = 17$. Kosmulski introduced this index arguing that less work is needed

in the calculation of $h^{(2)}$ in comparison with h but that is not really true. It is true that the h -index uses more ranks than the $h^{(2)}$ -index but h is directly visible from a ranked list of publications with citation scores while for $h^{(2)}$ one needs a comparison with the squared ranks. Anyway, we can say that both indices are easy to calculate but that the $h^{(2)}$ -index is even more insensitive to the actual numbers of citations per paper in a ranked list since Kosmulski's index uses much less papers than the h -index. We note however that Hua, Rousseau, Sun and Wan (2009) have found an interesting application of Kosmulski's $h^{(2)}$ -index in the framework of papers vs. the number of times these papers were downloaded (so where citations are replaced by downloads). Here Kosmulski's index uses more papers than in the citation case since download quantities are much higher than citation quantities.

Also this paper finds a new application of Kosmulski's $h^{(2)}$ -index. Indeed, with our method of truncating lists from above, and then continuing calculating the same index for the truncated list, more papers are involved. It is even so that, because $h^{(2)} \sqsupseteq h$, we have more "check points" in the ranked list of papers than with the h -index in the definition of CSS. Experimental results on this author's paper-citation scores will indicate the validity of the CSS method on the $h^{(2)}$ -index.

It is clear that the characteristic scores β_k (using the $h^{(2)}$ -index) are defined in the same way as for the h -index, where we now have $v_k = \sum_{i=0}^{k-1} h_i^{(2)}$ (and $v_1 = h^{(2)} =: h_0^{(2)}$) and $\gamma(v_k) = \beta_k$, yielding the CSS in this case.

Modeling this CSS is even more difficult than for the h -index. Also, there does not yet exist a similar result as in (6) (for $h = h_0$) for the $h^{(2)}$ -index. Hence also this will be done in the next proposition.

Proposition 3: Supposing the law of Lotka (1), we have that $h^{(2)} = h_0^{(2)}$ satisfies

$$h_0^{(2)} = h^{(2)} = T^{\frac{1}{2\alpha-1}} \quad (19)$$

$$h^{(2)} = h^{\frac{\alpha}{2\alpha-1}} \quad (20)$$

Hence for high α we have $h^{(2)} \approx \sqrt{h}$. The next value $h_1^{(2)}$ satisfies

$$h = \left(h_1^{(2)}\right)^{2\left(1-\frac{1}{\alpha}\right)} \left(h_1^{(2)} + h^{(2)}\right)^{\frac{1}{\alpha}} \quad (21)$$

$h_1^{(2)}$, only in terms of $h^{(2)}$ is the following relation

$$h^{(2)} = \left(h_1^{(2)}\right)^{1-\frac{1}{2\alpha-1}} \left(h_1^{(2)} + h^{(2)}\right)^{\frac{1}{2\alpha-1}} \quad (22)$$

which yields, for $\alpha = \frac{3}{2}$,

$$h_1^{(2)} = h^{(2)} \frac{\sqrt{5}-1}{2} \quad (23)$$

, the Golden Section of a line piece of length $h^{(2)}$.

Proof: The definition of the $h^{(2)}$ -index yields

$$\gamma\left(h^{(2)}\right) = \left(h^{(2)}\right)^2 \quad (24)$$

Hence, by (2), we have

$$B = \left(h^{(2)}\right)^{\beta+2}$$

Using (4) and (5) we get

$$\begin{aligned} h^{(2)} &= B^{\frac{1}{\beta+2}} \\ h^{(2)} &= \left(T^{\frac{1}{\alpha-1}}\right)^{\frac{1}{\beta+2}} \\ h^{(2)} &= T^{\frac{1}{2\alpha-1}} \end{aligned}$$

proving (19), from which (20) follows from (6). It is clear that

$$\lim_{\alpha \rightarrow \infty} h^{(2)} = \sqrt{h} \quad (25)$$

by (20). Now we will calculate $h_1^{(2)}$. The rank-frequency function for the truncated list is

$$\gamma_1(r) = \frac{B}{\left(r + h^{(2)}\right)^\beta} \quad (26)$$

$r \in [0, T - h^{(2)}]$. The definition of Kosmulski's index $h_1^{(2)}$ for this truncated list yields

$$\frac{B}{\left(h_1^{(2)} + h^{(2)}\right)^\beta} = \left(h_1^{(2)}\right)^2 \quad (27)$$

hence

$$B = \left(h_1^{(2)}\right)^2 \left(h_1^{(2)} + h^{(2)}\right)^\beta \quad (28)$$

Using (4), (5) and (6) yields $B = h^{1+\beta}$. Hence (28) yields (21).

By (20) and (5) B also equals

$$B = \left(h^{(2)}\right)^{\frac{2\alpha-1}{\alpha-1}}$$

hence, by (28) and (5), (22) follows. For $\alpha = \frac{3}{2}$, we hence have

$$h^{(2)} = \sqrt{h_1^{(2)} \left(h_1^{(2)} + h^{(2)}\right)}$$

from which (23) follows since $h_1^{(2)} > 0$ (same argument as in Proposition 1). \square

Note that the Golden Section is now obtained for $\alpha = \frac{3}{2}$ instead of $\alpha = 2$, in case of the h -index.

II.3 CSS for other h-type measures: R , g and $g^{(2)}$

The definition of the R -index – see Jin, Liang, Rousseau and Egghe (2007) – is as follows:

R is the square root of the total number of citations to the first h papers. Note that, for the definition of the R -index, the h -index is used. In terms of γ (formula (2)) we can write

$$R = \sqrt{\int_0^h \gamma(r) dr} \quad (29)$$

So CSS, for the R -index, are not easy to define for two reasons. First of all, the values of (29) are not always entire numbers; for cutting-off reasons we hence have to approximate R by the closest entire number. Secondly R_1 and h_1 (and the same for other indices) do not refer to the same truncated list. Therefore we do not use the R -index for the creation of CSS.

For the g -index, we can do better. The g -index has been defined in Egghe (2006 a,b,c) as the largest rank $r = g$ such the sum of the citations to the first g papers is at least g^2 . One could reformulate this as: the largest rank $r = g$ such that the first g papers have, on average, g citations. As proved in Egghe (2006a), for $\alpha > 2$, the g -index equals

$$g = \left(\frac{\alpha - 1}{\alpha - 2} \right)^{\frac{\alpha - 1}{\alpha}} h \quad (30)$$

with h as in (6).

The proof of (30) is given in Egghe (2006a) and is derived on the defining equation (using (2)):

$$\int_0^g \frac{B}{r^\beta} dr = g^2 \quad (31)$$

which yields (30), using (4) and (5). The same methodology will yield the g -index g_1 , of the truncated list, after cutting-off the first g papers. Now the rank-frequency function has the form (similar as for h and $h^{(2)}$):

$$\gamma_1(r) = \frac{B}{(r + g)^\beta} \quad (32)$$

with $r \in [0, T - g]$. The g -index g_1 for this truncated list is, by definition,

$$\int_0^{g_1} \frac{B}{(r + g)^\beta} dr = g_1^2 \quad (33)$$

Hence, since $0 < \beta < 1$ (since $\alpha > 2$),

$$\frac{B}{1 - \beta} \left[(g_1 + g)^{1 - \beta} - g^{1 - \beta} \right] = g_1^2 \quad (34)$$

which, implicitly, gives g_1 out of g and α and formulae (4), (5) and (6) (and (30)) (using these formulae one has that $B = h^{1 + \beta}$). As for the other indices, the CSS are determined by

$$\nu_k = \sum_{i=0}^{k-1} g_i \text{ and } \beta_k = \gamma(\nu_k), \text{ with } \gamma \text{ as in (2).}$$

A similar approach can be adopted for the $g^{(2)}$ -index. The $g^{(2)}$ -index has been introduced in Egghe (2009a) as the “Kosmulski-variant” of the g -index: the $g^{(2)}$ -index is the largest rank $r = g^{(2)}$ such that all papers on rank $1, \dots, g^{(2)}$ together have at least g^3 citations. In Egghe

(2009a) we did not present a formula for $g^{(2)}$ in the Lotkaian framework (analogous to (6) or (20)). We have the following proposition.

Proposition 4: Supposing the law of Lotka (1) with $\alpha > 2$, we have that $g^{(2)}$ satisfies

$$g^{(2)} = \left(\frac{\alpha - 1}{\alpha - 2} \right)^{\frac{\alpha-1}{2\alpha-1}} T^{\frac{1}{2\alpha-1}} \quad (35)$$

Hence

$$g^{(2)} = \left(\frac{\alpha - 1}{\alpha - 2} \right)^{\frac{\alpha-1}{2\alpha-1}} h^{(2)} \quad (36)$$

Proof: The definition of the $g^{(2)}$ -index gives, with γ as in (2):

$$\int_0^{g^{(2)}} \gamma(r) dr = \left(g^{(2)} \right)^3 \quad (37)$$

yielding, since $0 < \beta < 1$ (since $\alpha > 2$) that

$$g^{(2)} = \left(\frac{B}{1 - \beta} \right)^{\frac{1}{\beta+2}} \quad (38)$$

Using (4) and (5), we hence find (35) and, by (19), also (36). \square

Also here we see that, for α high, $g^{(2)} \approx \sqrt{g}$; this follows from (36), (30) and (25). Also indices $g_k^{(2)}$ can be determined via an argument similar as the one given for g_k above. We

can, again, define $\nu_k = \sum_{i=0}^{k-1} g_i^{(2)}$ and $\beta_k = \gamma(\nu_k)$, with γ as in (2).

III. CSS for the data of Egghe

The data of Table 1 in the Appendix yield the following CSS for the h -index: $h = h_0 = 17$,

$h_1 = 12$ (compare this with (8), in case $\alpha = 2$, yielding $h \frac{\sqrt{5}-1}{2} = 10.51$), $h_2 = 8$, $h_4 = 6$,

$h_5 = 6$, $h_6 = 5$, $h_7 = 4$ and $h_8 = 3$. Note that, indeed, $\frac{h_2}{h_1} > \frac{h_1}{h_0}$ which is agreeing with the

result in Proposition 2 in case $\alpha = 2$. The above values indicate that $\frac{h_k}{h_{k-1}}$ increases in k .

In each case above, the β_k can be determined via the rank $r = h_0 + h_1 + \dots + h_k$ and the corresponding value in Table 1 for the number of citations at that rank (e.g. for $\nu_1 = h_0 + h_1$, we have $\beta_1 = 12$).

The CSS from Table 1 for the g -index yields: $g = g_0 = 25$, $g_1 = 11$, $g_2 = 8$, $g_3 = 7$, $g_4 = 6$, $g_5 = 5$, $g_6 = 4$, $g_7 = 3$, again indicating that $\frac{g_k}{g_{k-1}}$ increases in k . In the same way as for the h -index we can generate from these values of g_k , the values ν_k and β_k .

For Kosmulski's $h^{(2)}$ -index, the values are only informative for lower values of k : $h^{(2)} = 6$, $h_1^{(2)} = 4$, $h_2^{(2)} = 4$, $h_3^{(2)} = 4$. Further $h_4^{(2)} = \dots = h_9^{(2)} = 3$, $h_{10}^{(2)} = \dots = h_{24}^{(2)} = 2$. Similar conclusions for the $g^{(2)}$ -index: $g^{(2)} = 6$, $g_1^{(2)} = 4$, $g_2^{(2)} = 4$, $g_3^{(2)} = 4$, $g_4^{(2)} = \dots = g_9^{(2)} = 3$, $g_{10}^{(2)} = \dots = g_{25}^{(2)} = 2$.

IV. Conclusions and suggestions for further research

In this paper we have noticed that the h -type indices only indicate one level of high-impact articles. The higher the value of the h -type-index, the more the need for other “marks” on the sequence of papers of an author, ranked in decreasing order of the number of received citations. The general principle (say for the h -index but similarly for other h -type-indices such as the g -index, the $h^{(2)}$ -index and the $g^{(2)}$ -index) is to determine $h = h_0$ = the h -index and then discard the papers in the h -core. The remaining list is renumbered so that the ranks again start at $r = 1$ and the h -index h_1 of this list is calculated. Again we discard these (second level) top h_1 articles and the remaining list is renumbered so that the ranks again start at $r = 1$ and the h -index h_2 of this list is calculated, and so on. The CSS are then, for

$k = 1, 2, \dots, \nu_k = \sum_{i=0}^{k-1} h_i$ and $\beta_k = \gamma(\nu_k)$ with γ the rank-frequency function of the original list.

A similar construction applies to other h -type indices.

These extra “marking points” better show the impact of the complete list of papers and in any case show the h -type performance of the papers of rank $h+1, h+2, \dots$, which are not used in the calculation of the h -index $h = h_0$ at all and which still have many citations (depending on h , which is high for prolific authors).

It is clear that more studies on CSS should be conducted and especially in the connection of h -type-indices. A challenge is to compare CSS of different authors, i.e. to determine how to compare such sequences (ν_k) and (β_k) for different authors.

We also leave open to determine the “requested” properties for CSS, i.e. what type of sequences illustrate best the impact (performance) of an author’s career.

References

- P. Ball (2005). Index aims for fair ranking of scientists. *Nature* 436, 900.
- L. Egghe (2005). *Power Laws in the Information Production Process: Lotkaian Informetrics*. Oxford, Elsevier.
- L. Egghe (2006a). Theory and practice of the g-index. *Scientometrics* 69(1), 131-152.
- L. Egghe (2006b). How to improve the h-index. *The Scientist* 20(3), 14.
- L. Egghe (2006c). An improvement of the h-index: the g-index. *ISSI Newsletter* 2(1), 8-9.
- L. Egghe (2009a). An econometric property of the g-index. Preprint.
- L. Egghe (2009b). The Hirsch-index and related impact measures. *Annual Review in Information Science and Technology*, to appear.
- L. Egghe (2009c). Characteristic scores and scales in a Lotkaian framework. Preprint.
- L. Egghe and R. Rousseau (2006). An informetric model for the Hirsch-index. *Scientometrics* 69(1), 121-129.

- W. Gellert, H. Küstner, M. Hellwich and H. Kästner (eds.) (1975). The VNR Concise Encyclopedia of Mathematics. Van Nostrand Reinhold Company, New York, USA.
- W. Glänzel (2008). What are your best papers? ISSI Newsletter 4(4), 64-67.
- W. Glänzel and A. Schubert (1988). Characteristic scores and scales in assessing citation impact. *Journal of Information Science* 14, 123-127.
- J. E. Hirsch (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(46), 16569-16572.
- P.-H. Hua, R. Rousseau, X.-K. Sun and J.-K. Wan (2009). A download $h^{(2)}$ -index as a meaningful indicator of Web impact of academic journals. Preprint.
- B. Jin, L. Liang, R. Rousseau and L. Egghe (2007). The R- and AR-indices: Complementing the h-index. *Chinese Science Bulletin* 52(6), 855-863.
- M. Kosmulski (2006). A new Hirsch-type index saves time and works equally well as the original h-index. *ISSI Newsletter* 2(3), 4-6.

Appendix

Table 1: Citation data of L. Egghe (based on WoS, February 4, 2009). 70 highest cited papers

r	#	r^2	$\sum\#$	r	#	r^2	$\sum\#$	r	#	r^2	$\sum\#$
1	58	1	58	24	13	576	620	47	8		
2	52	4	110	25	13	625	633	48	7		
3	52	9	162	26	13	676	646	49	7		
4	45	16	207	27	12			50	7		
5	41	25	248	28	12			51	7		
6	37	36	285	29	12			52	7		
7	27	49	312	30	11			53	6		
8	27	64	339	31	11			54	6		
9	25	81	364	32	11			55	6		
10	24	100	388	33	11			56	6		
11	20	121	408	34	10			57	6		
12	19	144	427	35	9			58	6		
13	19	169	446	36	9			59	5		
14	18	196	464	37	9			60	5		
15	18	225	482	38	9			61	5		
16	18	256	500	39	8			62	5		
17	17	289	517	40	8			63	5		
18	17	324	534	41	8			64	5		
19	16	361	550	42	8			65	4		
20	15	400	565	43	8			66	4		
21	15	441	580	44	8			67	4		
22	14	484	594	45	8			68	4		
23	13	529	607	46	8			69	4		
								70	3		

From Table 1 and the definition of the h -, g -, R -, $h^{(2)}$ - and $g^{(2)}$ -index, we have, clearly,

$$h=17, g=25, R=22,74, h^{(2)}=g^{(2)}=6.$$

To illustrate how we calculate h_1 , we cut-off the first $h=17$ papers and relabel the ranks of the remaining papers from $r=1$ onwards. This yields Table 2 (13 ranks suffice).

Table 2. Truncated and relabelled Table from the 18th paper onwards from Table 1.

r	#
1	17
2	16
3	15
4	15
5	14
6	13
7	13
8	13
9	13
10	12
11	12
12	12
13	11

It is clear from Table 2 that $h_1 = 12$. We finally illustrate the calculation of g_1 (the ones of $h_1^{(2)}$, $g_1^{(2)}$ and higher indices h_k , g_k , $h_k^{(2)}$, $g_k^{(2)}$, $k = 2, 3, \dots$ are similar).

Cut-off the first $g = 25$ papers and relabel the ranks of the remaining papers from $r = 1$ onwards. This yields Table 3 (12 ranks suffice).

Table 3. Truncated and relabelled Table from the 26th paper onwards from Table 1.

r	#	r^2	$\sum \#$
1	13	1	13
2	12	4	25
3	12	9	37
4	12	16	49
5	11	25	60
6	11	36	71
7	11	49	82
8	11	64	93
9	10	81	103
10	9	100	112
11	9	121	121
12	9	144	130

It is clear from Table 3 that $g_1 = 11$. Note that $h_1 > g_1$. This is not in contradiction with the fact that the g -index is always larger than or equal to the h -index. Indeed h_1 and g_1 are the h -

index, respectively g -index of different tables (because we truncate differently: for h_1 at rank 17 and for g_1 at rank 25).