# The *success*-index: an alternative approach to the *h*-index for evaluating an individual's research output

Fiorenzo Franceschini[1], Maurizio Galetto, Domenico Maisano, Luca Mastrogiacomo

[1] *fiorenzo.franceschini@polito.it*
Politecnico di Torino, DISPEA (Department of Production Systems and Business Economics),
Corso Duca degli Abruzzi 24, 10129, Torino (Italy)

**Abstract**

Among the most recent bibliometric indicators for normalizing the differences among fields of science in terms of citation behaviour, Kosmulski (Journal of Informetrics 5(3):481–485, 2011) proposed the *NSP* (Number of Successful Paper) index. According to the authors, *NSP* deserves much attention for its great simplicity and immediate meaning – equivalent to those of the *h*-index – while it has the disadvantage of being prone to manipulation and not very efficient in terms of statistical significance.

In the first part of the paper, we introduce the *success*-index, aimed at reducing the *NSP*-index's limitations, although requiring more computing effort. Next, we present a detailed analysis of the *success*-index from the point of view of its operational properties and a comparison with the *h*-index's ones. Particularly interesting is the examination of the *success*-index scale of measurement, which is much richer than the *h*-index's. This makes *success*-index much more versatile for different types of analysis – e.g., (cross-field) comparisons of the scientific output of (1) individual researchers, (2) researchers with different seniority, (3) research institutions of different size, (4) scientific journals, etc..

**Keywords:** Successful paper, *NSP*-index, Field normalization, Reference practices, Operational properties, Hirsch index.

## 1. Introduction

Defining bibliometric impact measures, which allow cross-field normalization without being influenced by the different propensity to cite, is an age-old problem still much debated [Garfield, 1979a]. Many indicators, mostly related to scientific journals, have been proposed over the years. Not surprisingly, a common feature of these indicators is that they are based on the comparison between (1) the amount of citations received by a group of publications examined and (2) a *comparison term* given by the expected number (or another indicator of central tendency) of the citations received or made by analogous publications in the specific discipline(s) of interest. At the risk of oversimplifying, the cross-field normalized impact indicators proposed in the literature differ in three main features:

1. A first distinction is about how the comparison term is determined. In the so-called *target* or *cited-side* normalization, this term is given by the number of citations received, on average, by a reference sample of publications within the discipline(s) of interest. In the so-called *source* or *citing-side* normalization, it is obtained using the average number of citations made (i.e., bibliographic references) by a reference sample of publications, which is a widely accepted estimator of the "field's propensity to cite" [Moed, 2010a; 2010b].
2. A second distinction concerns the "moment" in which the normalization is performed. A first option is to build an indicator based on the citations received by a group of reference publications and subsequently normalize it (*a posteriori* normalization). Another solution is to immediately normalize citations before joining them by an aggregated indicator (*a priori* normalization or *fractional counting*) [Pinski and Narin, 1976; Zitt and Small, 2008; Leydesdorff and Opthof, 2010; Glänzel et al., 2011].
3. Another distinctive feature is the selection of a reference sample of publications for determining the comparison term, based on the citations received or made by the selected publications. This is probably one of the trickiest and most controversial aspects. Some techniques are *classification-dependent*, since they are based on a superimposed delineation of fields of science, e.g., that one provided by ISI-Thomson Reuters or other bibliometric databases [Moed, 2010b; Leydesdorff and Shin, 2011]. Other techniques are more "adaptive", since the sample is determined considering the "neighbourhood" of the publication(s) of interest – typically consisting of the set of publications citing or being cited by them [Jackson and Rogers, 2007; Waltman et al. 2011a; Waltman et al. 2011b].

In addition, the normalization approaches can be differentiated by secondary aspects, such as (1) the size of the time-window for counting the citations received (or made) by the publications examined, (2) the way of calculating the comparison term, (3) the way of determining the neighbourhood of a publication (e.g., given by the "neighbour" publications only or by the "neighbours of the neighbours", etc.).

Despite the apparent differentiation among the approaches to normalization, we believe that – if examined carefully – their structure is not so dissimilar. For example, even the most *sui generis* normalised impact measures, such as those based on the fractional counting, can be reduced to the ratio between the number of citations received by one or more publications of interest and a comparison term represented by a central tendency indicator (specifically, the harmonic mean) associated to the citations received or made by a sample of "external" publications within the field concerned [Small and Sweeney, 1985; Zitt and Small, 2008; Guns and Rousseau 2010; Zhou and Leydesdorff, 2011].

Most of the normalised indicators have been historically defined at journal level, with the aim of enabling cross-field comparisons [Garfield, 1979b]. However, with appropriate adaptations, they can be applied to the publications of individual scientists, enabling comparisons among scientists from different disciplines.

It is worth noting that the cross-field normalization can also apply at level of a scientist's individual publications. It is not so rare for a scientist to be involved in research topics at the boundary between fields having different propensities to cite. A plethora of examples can be found: e.g., Energetics and Chemistry, Biomedical Engineering and Medicine, Social Sciences and Statistics, etc..

Despite its relative recentness, the *h*-index is probably the most in vogue among the indicators used for evaluating the performance of individual scientists [Hirsch, 2005]. One of the most important merits behind its success and popularity is the great simplicity and the immediate intuitive meaning [Franceschini and Maisano, 2010a]. Whereas, an important drawback is that it does not allow comparisons among scientists of different disciplines. Several attempts to resolve this limitation, like those of Iglesias and Pecharromán (2007) and Batista et al. (2006), have been made, but the original simplicity and immediacy of *h* is often undermined by the corrections introduced for its normalization [Franceschini and Maisano, 2010a; 2010c; 2011]. Moreover, the proposed normalizations are based on the assumption that the whole scientific production of a scientist is "homogeneous", i.e., all the publications are included within the same discipline.

In a recent paper in *Journal of Informetrics*, Kosmulski (2011) introduced a new interesting indicator called *Number of Successful Paper* (*NSP*). Similarly to *h*, *NSP* makes it possible to "isolate" a subset of publications, defined as "successful papers", among a group of publications examined – e.g., those associated to a scientist or a journal. Precisely, a publication is classified as successful when it has received more citations than those made.

In other words, a score is associated to each (*i*-th) of the (*P*) publications of interest:

$$\begin{aligned} score_i &= 1 \qquad \text{when } c_i > r_i \\ score_i &= 0 \qquad \text{otherwise} \end{aligned} \tag{1}$$

where $c_i$ are the citations received and $r_i$ the citations made by the *i*-th publication.

*NSP* is defined as:

$$NSP = \sum_{i=1}^{P} score_i \ . \tag{2}$$

What is most fascinating about *NSP* is the simplicity and immediateness of meaning, comparable to those of the *h*-index.

The purpose of this paper is to make explicit some reflections on the Kosmulski's *NSP*-index, whose structure is particularly interesting among the current normalized impact measures. The authors show that *NSP* appears superior to *h* in terms of some operational properties and has the potential to replace it in certain contexts.

The remaining of this paper is organised as follows. Section 2 contains a short digression on the practice of publishing and citing in the scientific literature. This will be helpful in understanding the similarities and potential limitations of the normalized impact measures in general, and to prepare the field for the subsequent discussion on *NSP*. Section 3 reports a critical analysis of the *NSP*-index, suggesting some modifications and a new variant (the *success*-index). Section 4 proposes a structured comparison between the *success*-index and the *h*-index, on the basis of their major operational properties, focusing on the relevant pros and cons.

Finally, the conclusions are given, emphasizing the potentialities of the *success*-index and summarising the original contribution of the paper.

## 2. Different disciplines with different propensities to cite

Assuming that we look into the whole scientific literature, including the totality of the journal publications, monographs, book chapters, conference proceedings and other types of publications ever issued, we may represent this "universe" – from the citational point of view – as a graph formed by a dense network of interconnected nodes (see the simplified representation of Fig. 1). Each ($i$-th) node represents a scientific paper ($P_i$), while arcs represent the relevant citations. Precisely, incoming arrows are the ($c_i$) citations received by other papers and outgoing arrows are the ($r_i$) citations made (or bibliographic references) to other papers. From the viewpoint of a generic node/paper, the practice of citing can be therefore seen actively or passively.

The meaning of the citation, as an expression of interest in a (cited) publication by another (citing) publication has been debated in the literature for many decades [Small, 2004]. Citation is traditionally considered in its passive acceptation as a credit or acknowledgement received from other publications in the literature. In this sense, the mostly cited publications are likely to be those of greater impact.

For the determination of a comparison term that represents the propensity to cite, there is no apparent difference between using the average citations *received* or *made* by a sample of scientific papers. The following example clarifies this aspect.

Assuming that the totality of the ($P$) publications of the scientific literature are captured by means of an ideal "omni-covering" bibliometric database – i.e., able to index all the possible scientific publications – and assuming that no citations from these publications are directed to "external resources" – i.e., non-scientific publications, such as patents, websites, etc...), the total citations ($C$) will perfectly balance the total references ($R$) (see Eq. 3). In fact, they are exactly the same thing, i.e., the arcs of the graph.

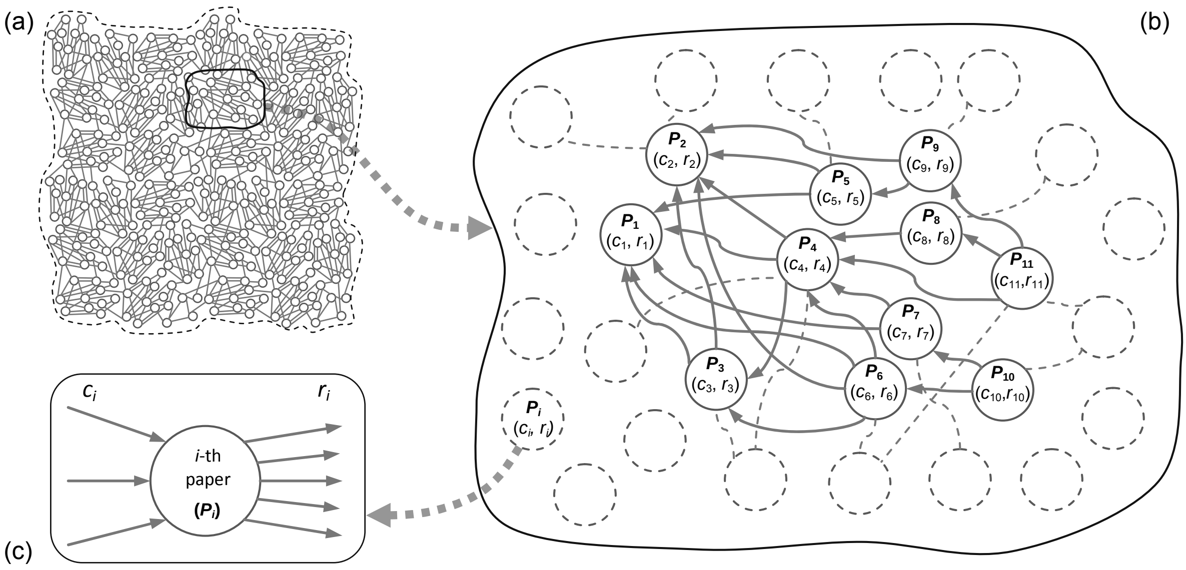


**Fig. 1. (a) Representation of the scientific literature by a graph. (b) Nodes represent scientific papers while arcs represent mutual citations. (c) Regarding the *i*-th paper ($P_i$), $c_i$ denotes the total citations received (incoming arrows), while $r_i$ denotes the total citations made (outgoing arrows).**

$$C = \sum_{i=1}^{P} c_i \quad = \quad R = \sum_{i=1}^{P} r_i \ . \tag{3}$$

Obviously, the arithmetic mean of the citations ($\overline{c}$) and references per publication ($\overline{r}$) will coincide:

$$\overline{c} = \frac{1}{P} \cdot \sum_{i=1}^{P} c_i \quad = \quad \overline{r} = \frac{1}{P} \cdot \sum_{i=1}^{P} r_i \ . \tag{4}$$

We remark that the equality in Eq. 4 is not necessarily satisfied for other indicators of central tendency, different from the arithmetic mean (e.g., median, geometric mean, harmonic mean, etc.).

A first observation is that the two previous indicators ($\overline{c}$ and $\overline{r}$) are not practically useful, since they are calculated mixing together publications of disparate disciplines, in which – it is well-known – the propensity to cite can be very different [JCQAR, 2010].

A second observation is that the variability of the $c_i$ values is likely to be higher than that of $r_i$ values [Ravichandra Rao, 2011]. While $r_i$ values cannot be said to be distributed uniformly, their distribution is probably less uneven than that one of the $c_i$ values (which, by its nature, is generally very skewed). This statement can be confirmed by several studies on uncitedness – for example [Larivière et al., 2008] – in which it appears that a large percentage of scientific publications (very often larger than 50%) are not cited although they – almost certainly – cite other publications.

A third observation is that, for a real bibliometric database with (inevitably) limited coverage, the sum of $r_i$ values over all scientific publications is certainly higher than the sum of their $c_i$ values, due to references to non-scientific publications or to scientific publications not covered.

Going back to the ideal representation and assuming that an *a priori* field delineation of the whole set of publications is performed – i.e., publications are divided into groups associated with the various disciplines – there will certainly be a significant difference between groups in terms of average propensity to cite. Obviously, the use of indicators influenced by the different propensity to cite – as the *h*-index – should be limited to evaluations within each group, in order to avoid “comparisons between apples and oranges”.

Also, there will not necessarily be a perfect balance between the total number of citations received ($C_g$) and made ($R_g$), regarding the publications of a certain group. The same applies to the respective mean values ($\bar{c}_g$ and $\bar{r}_g$, see Fig. 2). The (obvious) reason is that groups of publications are not perfectly “isolated” but may exchange citations [Zitt, 2010]. In general, basic or “interdisciplinary” scientific disciplines may be likely to attract more citations from "external" publications. The effect of these exchanges will be even more pronounced when reducing the perimeter of the publications examined, for example when passing from disciplines to sub-disciplines or to other sub-groups of publications (e.g., those contained in one or more journals). Therefore, Eqs. 3 and 4 will not necessarily be satisfied when limiting the domain to the publications of a specific (sub-)discipline.

In general, the more the exchanges with the exterior are, the more $\bar{c}_g$ will be likely to deviate from $\bar{r}_g$. For a group of publications, $\bar{r}_g$ can be seen as a rough estimator of $\bar{c}_g$, but this estimation is likely to be better when the citation exchange with external publications is little. For more detailed and rigorous information on the relation between the $\bar{r}_g$ and $\bar{c}_g$ values concerning a set of documents, we refer the reader to [Egghe and Rousseau, 1990; Kochen, 1974].

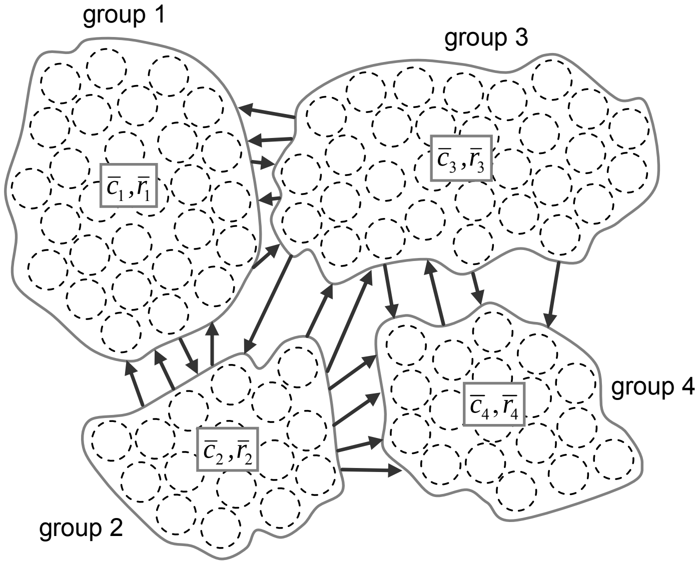


**Fig. 2. Schematic representation of subject-based groups (represented by solid closed lines) of papers (represented by dotted circles) in the scientific literature. Each group has a peculiar propensity to cite, which can be estimated by the mean citations or mean references per paper ($\bar{c}_g$ and $\bar{r}_g$), or other indicators. Arrows represent the citations exchanged between papers of different groups. Some disciplinary groups (such as group 4) have a greater tendency to receive citations than others (this might be the case of basic or “interdisciplinary” scientific disciplines) and vice versa.**

A typical way of estimating $\bar{c}_g$ and $\bar{r}_g$ within a discipline is to use a reference sample (or subset) of

publications. While small samples tend to bring to a poor estimation, the risk of using too large samples is to mix together publications from different (sub-)disciplines and to make an estimation that does not reflect the propensity to cite within the specific area of interest.
It is now clear why the selection of a reference sample of publications is crucial when estimating the comparison term for normalization. Among the most recent techniques in the literature, it would seem that those based on "neighbourhood" tend to prevail over those based on the *a priori* delineation of the scientific fields. The reasons may be various:

1. First, it is practically impossible to determine an objective classification of scientific journals in "macro-disciplines", because their boundaries are fuzzy. Even assuming that it was, there is often a great fragmentation in sub-disciplines, with important differences in terms of propensity to cite [Garfield, 1979a].
2. The classification of scientific journals in pre-established disciplines, as well as being arbitrary, is incomplete because it is limited to a subset of journals indexed by the bibliometric database of interest (e.g., ISI Web of Science (WoS), Google Scholar (GS) or Scopus), whose coverage – in some areas especially – is far from being complete [Harzing and van der Wal, 2008].
3. On the other hand, estimating the propensity to cite by a sample of publications that represent the "neighbourhood" (provided that it is not too small, say not smaller than 20-30 publications) seems to be a more "adaptive" and accurate method.

Assuming that identifying a reference sample of publications is not an issue, what is the most appropriate indicator for the estimation of the comparison term? E.g., mean/median citations *received* or mean/median citations *made* by the sample publications? The first probably appears as more direct. On the other hand, the second may lead to several advantages:

- The number of citations made (related to a reference sample of publications) is fixed over time, while the number of citations received tends to increase and requires a physiological accumulation period to stabilize – typically, around 3-5 years depending on the disciplines [Amin and Mabe, 2000]. For this reason, indicators based on the number of references look more stable and robust, especially for relatively recent samples of publications.
- This stability is also derived by the fact that the number of references is likely to be, on average, less variable than the number of citations received. The estimation will therefore be less subject to fluctuations.
- Certainly, the citations that a *present* publication will receive will come from *future* publications. Therefore, it is somehow questionable to estimate the future propensity to cite by the present one. However, since changes in the propensity to cite generally require a large number of years (hardly less than 10-15 years [Bornmann, 2011; Kranpen, 2010]), the result of

this approximation is not very distorted.

### 3. Reflections on the Kosmulski's *NSP*-index and introduction of the *success*-index

The structure of the *NSP*-index is significantly different from that of other normalized impact indicators. Like the *h*-index, it isolates a subset of more relevant publications from the rest of the production.

A questionable feature of *NSP*, as recognized by Kosmulski (2011), is that the "threshold" for "promoting" or "rejecting" a publication is given by a certain number of citations and no account is taken of the "excess citations" [Franceschini and Maisano, 2010c]. The same effect is present for the *h*-index. Indeed every threshold is somewhat arbitrary and its use can be questionable, although we believe that – regarding *NSP* – this will be compensated by its simplicity and immediate meaning.

A potentially more controversial aspect of the *NSP*-index is the construction of the comparison term (or threshold): in this case the number of citations made (i.e., the number of references) by a publication. According to the authors, this construction may be questionable for two main reasons:

1. As known from Statistics, the smaller the sample population, the larger the variability/uncertainty associated with the estimation of a parameter. Therefore, estimations based on unitary samples are the crudest.
2. While a scientist may only marginally affect the citations received by his/her publications (typically by self-citations, etc.), on the other hand the number of references can be influenced very easily. A scientist, for having the largest possible *NSP*s, may be tempted to "economize" on the list of references, with the risk of depriving the reader of some useful information for better understanding the details and nuances of his/her research [Haslam et al., 2005; Ravichandra Rao, 2011]. It must be said that this manipulation is partially impeded by the fact that (a) referees and editors can (and very often do) force the author(s) to add relevant references, and (b) it has no immediate effect since it will take some time (to accumulate citations) before a new paper contributes to *NSP*. Of course, we recognize that no indicator (bibliometric and otherwise) is totally immune from manipulation [Franceschini et al., 2007]; in the case of *NSP*, however, the problem seems particularly manifest.

For completeness, we must not forget the main advantages of the term of comparison, as originally defined by Kosmulski:

- Apart from deterring self-citations, this indicator would discourage "heavy citers" – i.e., those scientists who cite "too easily", just to give more credit to their argument – although the task of "cleaning" the unnecessary citations should probably be carried out by referees [Williamson, 2009].

- In case of special papers, such as reviews, there is a “self-compensation” between the high number of citations that they are likely to receive and the as high number of references that they generally have. However, it is worth remarking that the main bibliometric databases (such as WoS or Scopus) make it possible to separate reviews from research articles easily.
- Being based on data that are immediately available from the main bibliometric databases (i.e., the $c_i$ and $r_i$ values associated to a publication of interest), the calculation of *NSP* is relatively simple and fast.

Apart from the benefits above, the greatest weakness of *NSP* lies in the dubious estimation of the comparison term. To further clarify this, let us consider the example in Tab. 1, representing the number of citations received ($c_i$) and made ($r_i$) by the publications issued by four scientific journals (i.e., Monatshefte fur Chemie (MC), Monatshefte fur Mathematik (MM), Journal of Neural Transmission (NT) and Chemical Reviews (CR)) in the year 2000. Articles were analyzed by means of the Scopus database, accessed on June, 2011. We decided to refer to these journals since they had been used by Kosmulski (2011) to exemplify the use of *NSP*-index. Even though every journal reasonably contains articles within the same discipline – and therefore with (roughly) the same propensity to cite – it can be noticed that the variability relating to the number of references per publication is relatively high. This is an empirical proof that the estimation of the comparison term through a single paper is very questionable.

Also, the relationship between $c_i$ and $r_i$, at the level of individual publication, seems to be very weak, as denoted by the low slopes and $R^2$ values of the tendency lines. Thus, the hypothesis that publications with more references have a certain advantage in collecting more citations does not seem to be respected at least in this case [Davis, 2010].

In our opinion, this kind of conceptual mistake is somehow similar to a classical misuse of the Impact Factor (*IF*): the *IF* – i.e., the average number of citations received by a journal in a certain time-window – is often used (wrongly) as an indicator of impact relating to a single article [Amin and Mabe, 2000; Garfield, 2005]. As regards the *NSP*-index, the $r_i$ value of a single publication is used, alone, as an estimator of the average number of references made by the publications of a certain (sub)discipline.

To avoid this problem, we propose a *success*-index using exactly the same formulation of Kosmulski, with the only exception that – for each $i$-th publication – the comparison term is not necessarily given by $r_i$, but is replaced by a more appropriate indicator of propensity to cite, determined on the basis of a representative sample of publications. Similarly to *NSP*, a score is associated to each ($i$-th) of the ($P$) publications of interest:

$$
\begin{aligned}
&score_i = 1 \quad &\text{when } c_i > CT_i \\
&score_i = 0 \quad &\text{otherwise}
\end{aligned}
\tag{5}
$$

where $CT_i$ is a generic comparison term associated to the *i*-th publication. We underline that $CT_i$ is an estimate of the number of citations that a publication – in a certain scientific context and period of time – should potentially achieve. This achievement determines the condition of success. In this sense, $CT_i$ embodies the concept of *potential citation impact* by Moed (2010a).

Similarly to Eq. 2, the *success*-index is given by:

$$
Success\text{-index} = \sum_{i=1}^{P} score_i \ . \tag{6}
$$

With reference to the general considerations in Sect. 2, possible ways of constructing $CT_i$ are:

- $r_i$, i.e. the number of citations made by the (*i*-th) publication concerned (case of the Kosmulski's *NSP*-index);
- $(\bar{r}_{JY})_i$ or $(\tilde{r}_{JY})_i$, i.e. the mean or median number of references made by the articles published in the same journal (*J*) and year (*Y*) of the (*i*-th) publication concerned;
- $(\bar{c}_{JY})_i$ or $(\tilde{c}_{JY})_i$, i.e. the mean or median number of citations received by the articles published in the same journal (*J*) and year (*Y*) of the (*i*-th) publication concerned;
- $(\bar{r}_N)_i$ or $(\tilde{r}_N)_i$, i.e. the mean or median number of references made by a sample of publications representing the "neighbourhood" of the (*i*-th) publication concerned;
- $(\bar{c}_N)_i$ or $(\tilde{c}_N)_i$, i.e. the mean or median number of citations received by a sample of publications representing the "neighbourhood" of the (*i*-th) publication concerned.

All these $CT_i$s represent possible estimates of the citations that a new publication would be likely to receive in the field of interest. A potential problem when constructing $CT_i$ referring to the articles published by a certain journal (e.g., in the case of $(\tilde{r}_{JY})_i$ or $(\tilde{c}_{JY})_i$) is to mix up different types of papers (e.g., research articles, notes, letters to the editor, reviews, etc..), with different tendency to make/receive citations. The problem can be generally overcome by "filtering out" certain types of papers.

Undoubtedly, the calculation of the *success*-index is more complicate than that of *NSP*, because – for each publication – it requires the selection of a sample of "external" publications for determining the comparison term. However, we believe that this correction is essential for reducing the indicator's manipulability and increasing its effectiveness in terms of statistical significance. Note that we have not put any constraint on the definition of the new comparison term, provided that it must be based on a reasonably representative sample of publications, "close" to that one of interest. As a consequence, the typical issues concerning (1) the sample selection and (2) the choice of a suitable indicator for denoting the propensity to cite remain still open. It is worth mentioning that a more recent approach, evading the use of measures of central tendency, is the *I3* indicator as proposed by Leydesdorff and Bornmann (2011).

Despite the claimed "freedom" in the construction of $CT_i$, for the purpose of simplicity and practicality, it will be hereafter calculated as $\left(\tilde{r}_{JY}\right)_i$. This conventional choice can be justified by three reasons:

1. It seems reasonable to assume that the papers published by a journal in a given year are relatively homogeneous in terms of propensity to cite (except for "wide-ranging" journals, such as Nature or Science).
2. The calculation of this indicator is simpler than others, such as those based on "neighbour" publications. In a sense, the articles published by a journal in a given year will be considered as "part of the same family".
3. The median is more robust than the mean value, because it is less influenced by "outliers", that is publications with a number of citations made/received much higher/lower than the others [Bornmann and de Moya Anegon, 2011].

### *3.1 Empirical application examples of the success-index*

For the purpose of example, we now present the calculation of the *success*-index for the four journals seen before (see results in Tab. 1). Fig. 3 graphically represents the results relating to one of these journals (i.e. MM). We point out that the difference between the *NSP*- and the *success*-index values is due to the fact that the corresponding comparison terms are constructed according to two different logics.

In the next example, we propose the calculation of the *success*-index for two individual anonymous scientists: *Scientist 1* (*S1*), a chemist, and *Scientist 2* (*S2*), a computer scientist. The database used is Scopus and the articles selected are those published in the 2000-2005 time window, in order to make the comparison between scientists more homogeneous. The authors are aware that the size of the time window may importantly influence the result of the comparison; e.g., a too narrow time window may "destabilize" result's robustness. To avoid interference from editorial materials (prefaces, corrections), articles with $r_i = 0$ were deliberately ignored.

Results are illustrated in Tab. 2. As expected, the variability of the $r_i$ values is significantly higher than that relating to the $\left(\bar{r}_{JY}\right)_i$ and $\left(\tilde{r}_{JY}\right)_i$ values (see the corresponding standard deviations at the bottom of the table). By the way, it is interesting to observe that the $\left(\bar{r}_{JY}\right)_i$ values are always higher than the corresponding $\left(\tilde{r}_{JY}\right)_i$ values. Moreover, the correlation between $c_i$ and $r_i$ values is very weak ($R^2 \leq 0.07$ for both the scientists).

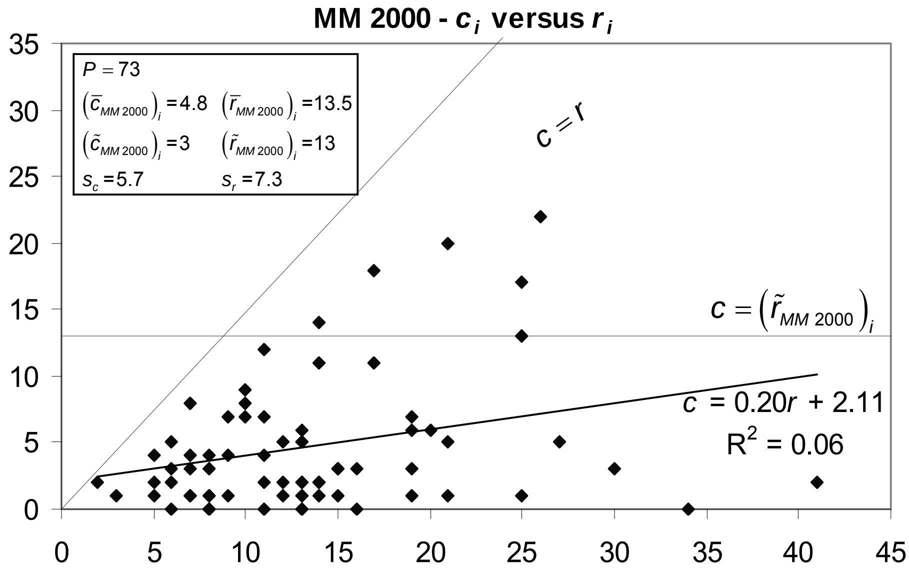


**Fig. 3. Diagram representing the relationship between $c_i$ and $r_i$ values, at the level of individual publications. The articles examined were issued by a journal (i.e., MM) in 2000. Articles have been analyzed by means of the Scopus database, accessed on June 2011. The total number of articles (*P*), the mean, the median value and the standard deviation relating to $c_i$ and $r_i$ values (respectively $(\bar{c}_{JY})_i, (\tilde{c}_{JY})_i, s_c, (\bar{r}_{JY})_i, (\tilde{r}_{JY})_i$ and $s_r$ ) are reported in the top-left square. The bisector (*c*=*r*) and the horizontal line *c*=$(\tilde{r}_{JY})_i$, being $(\tilde{r}_{JY})_i$ the median number of references associated to the articles of each journal. According to the definition of Kosmulski, the successful papers are the points falling above the bisector, whereas according to a possible formulation of the *success*-index (see the end of Sect. 3), they are the points above the horizontal line.**

| *Journal* | *P* | $(\bar{c}_{JY})_i$ | $(\tilde{c}_{JY})_i$ | $s_c$ | $(\bar{r}_{JY})_i$ | $(\tilde{r}_{JY})_i$ | $s_r$ | $R^2$ | *NSP* | *success*-index |
|---|---|---|---|---|---|---|---|---|---|---|
| MC | 148 | 8.6 | 6 | 9.4 | 24.3 | 21 | 15.5 | 0.23 | 13 | 13 |
| MM | 73 | 4.8 | 3 | 5.7 | 13.5 | 13 | 7.3 | 0.06 | 4 | 6 |
| NT | 123 | 26.1 | 16 | 27.0 | 38.7 | 31 | 27.9 | 0.34 | 26 | 34 |
| CR | 190 | 145.9 | 58.5 | 352.2 | 188.8 | 187.5 | 121.8 | 0.08 | 75 | 71 |

**Tab. 1. Example of calculation of the *NSP*- and the *success*-indices, considering the articles issued by four journals (i.e., MC, MM, NT and CR) in 2000. For the calculation of the *success*-index, $CT_i = (\tilde{r}_{JY})_i$ . Articles were analyzed by means of the Scopus database, accessed on June 2011. For each journal, the total number of articles (*P*), the mean, median, standard deviation relating to $c_i$ and $r_i$ values, and the $R^2$ values of the (1st order) relationship between the latters (respectively $(\bar{c}_{JY})_i, (\tilde{c}_{JY})_i, s_c, (\bar{r}_{JY})_i, (\tilde{r}_{JY})_i, s_r$, and $R^2$ ) are also reported.**

It can be noticed that the two scientists are comparable as regards overall productivity (37 versus 42 publications). Despite *S1* is better than *S2* in terms of total citations (*C*) and *h*-index, he/she is weaker in terms of *NSP*- and *success*-index. This is due to the generally higher propensity to cite in the scientific field of *S1*, compared to the *S2*'s (e.g. see the mean or median values of the relevant $CT_i$s, i.e. $r_i$ and $(\tilde{r}_{JY})_i$). In particular, *S2* has a *success*-index of 7 because 7 of his/her publications have a number of citations larger than the respective $(\tilde{r}_{JY})_i$ values, representing an estimate of the citations that a new publication would be likely to receive (on average) in the field of interest.

We also emphasize that the *success*-index, in contrast to *C* and *h*, is insensitive to the different propensity to cite, and therefore is suitable for comparisons between authors of different disciplines.

| *S1 (anonymous Chemist)* | | | | | | | |
|---|---|---|---|---|---|---|---|
| *Paper No.* | *Year* | $c_i$ | $r_i$ | $\left(\bar{r}_{JY}\right)_i$ | $\left(\tilde{r}_{JY}\right)_i$ | *NSP-core* | *success-core* |
| 1 | 2002 | 117 | 11 | 29.8 | 27.0 | ✓ | ✓ |
| 2 | 2004 | 109 | 13 | 23.0 | 20.0 | ✓ | ✓ |
| 3 | 2004 | 106 | 15 | 29.9 | 28.0 | ✓ | ✓ |
| 4 | 2002 | 101 | 153 | 78.0 | 58.5 | ✗ | ✓ |
| 5 | 2000 | 52 | 9 | 24.8 | 22.0 | ✓ | ✓ |
| 6 | 2003 | 43 | 19 | 28.6 | 26.0 | ✓ | ✓ |
| 7 | 2003 | 32 | 181 | 64.7 | 43.0 | ✗ | ✗ |
| 8 | 2004 | 30 | 59 | 65.0 | 52.5 | ✗ | ✗ |
| 9 | 2004 | 28 | 79 | 29.9 | 28.0 | ✗ | ✗ |
| 10 | 2002 | 21 | 24 | 34.5 | 29.0 | ✗ | ✗ |
| 11 | 2004 | 20 | 54 | 37.5 | 27.0 | ✗ | ✗ |
| 12 | 2002 | 19 | 13 | 41.1 | 37.0 | ✓ | ✗ |
| 13 | 2002 | 18 | 25 | 29.8 | 27.0 | ✗ | ✗ |
| 14 | 2003 | 16 | 17 | 29.4 | 28.0 | ✗ | ✗ |
| 15 | 2002 | 15 | 11 | 35.9 | 26.0 | ✓ | ✗ |
| 16 | 2000 | 15 | 16 | 29.6 | 26.0 | ✗ | ✗ |
| 17 | 2001 | 11 | 11 | 28.4 | 26.0 | ✗ | ✗ |
| 18 | 2002 | 10 | 15 | 29.8 | 27.0 | ✗ | ✗ |
| 19 | 2002 | 9 | 25 | 21.0 | 14.0 | ✗ | ✗ |
| 20 | 2003 | 8 | 10 | 56.4 | 53.0 | ✗ | ✗ |
| 21 | 2001 | 7 | 12 | 28.4 | 26.0 | ✗ | ✗ |
| 22 | 2005 | 6 | 7 | 30.9 | 29.0 | ✗ | ✗ |
| 23 | 2005 | 6 | 20 | 38.6 | 27.0 | ✗ | ✗ |
| 24 | 2002 | 5 | 30 | 21.0 | 14.0 | ✗ | ✗ |
| 25 | 2003 | 5 | 12 | 30.0 | 27.0 | ✗ | ✗ |
| 26 | 2001 | 4 | 6 | 18.1 | 15.0 | ✗ | ✗ |
| 27 | 2004 | 4 | 5 | 29.9 | 28.0 | ✗ | ✗ |
| 28 | 2003 | 3 | 9 | 28.6 | 26.0 | ✗ | ✗ |
| 29 | 2004 | 3 | 33 | 28.4 | 27.0 | ✗ | ✗ |
| 30 | 2004 | 3 | 15 | 29.9 | 28.0 | ✗ | ✗ |
| 31 | 2000 | 2 | 13 | 21.6 | 18.0 | ✗ | ✗ |
| 32 | 2003 | 2 | 9 | 28.6 | 26.0 | ✗ | ✗ |
| 33 | 2001 | 1 | 26 | 26.0 | 25.0 | ✗ | ✗ |
| 34 | 2002 | 1 | 71 | 33.7 | 22.0 | ✗ | ✗ |
| 35 | 2005 | 1 | 10 | 29.3 | 27.0 | ✗ | ✗ |
| 36 | 2004 | 0 | 3 | 29.3 | 22.0 | ✗ | ✗ |
| ***P*=37** | 2005 | 0 | 8 | 29.3 | 27.0 | ✗ | ✗ |
| Total | - | ***C*=833** | 1049 | - | - | | ***h*-index=15** |
| Mean | - | 22.5 | 28.4 | 33.2 | 28.1 | | ***NSP*-index=7** |
| Median | - | 9.0 | 15.0 | 29.8 | 27.0 | | ***success*-index=6** |
| St.dev. | - | 32.6 | 38.1 | 12.7 | 9.6 | | |

| *S2 (anonymous Computer Scientist)* | | | | | | | |
|---|---|---|---|---|---|---|---|
| *Paper No.* | *Year* | $c_i$ | $r_i$ | $\left(\bar{r}_{JY}\right)_i$ | $\left(\tilde{r}_{JY}\right)_i$ | *NSP-core* | *success-core* |
| 1 | 2000 | 53 | 30 | 22.1 | 18.0 | ✓ | ✓ |
| 2 | 2000 | 46 | 14 | 22.1 | 17.0 | ✓ | ✓ |
| 3 | 2000 | 22 | 42 | 15.7 | 12.0 | ✗ | ✓ |
| 4 | 2002 | 21 | 44 | 20.4 | 17.0 | ✗ | ✓ |
| 5 | 2000 | 20 | 24 | 22.1 | 17.0 | ✗ | ✓ |
| 6 | 2000 | 19 | 10 | 22.1 | 17.0 | ✓ | ✓ |
| 7 | 2005 | 19 | 14 | 40.1 | 38.0 | ✓ | ✗ |
| 8 | 2003 | 16 | 14 | 31.8 | 25.0 | ✓ | ✗ |
| 9 | 2002 | 14 | 20 | 28.5 | 27.5 | ✗ | ✗ |
| 10 | 2003 | 14 | 10 | 26.3 | 23.0 | ✓ | ✗ |
| 11 | 2005 | 14 | 13 | 33.3 | 31.0 | ✓ | ✗ |
| 12 | 2001 | 13 | 15 | 20.3 | 15.0 | ✗ | ✗ |
| 13 | 2005 | 13 | 12 | 40.1 | 38.0 | ✓ | ✗ |
| 14 | 2000 | 13 | 21 | 15.7 | 12.0 | ✗ | ✓ |
| 15 | 2004 | 11 | 18 | 23.6 | 19.0 | ✗ | ✗ |
| 16 | 2001 | 9 | 6 | 21.7 | 16.0 | ✓ | ✗ |
| 17 | 2004 | 9 | 17 | 29.0 | 25.5 | ✗ | ✗ |
| 18 | 2004 | 9 | 7 | 23.6 | 19.0 | ✓ | ✗ |
| 19 | 2005 | 9 | 14 | 40.1 | 38.0 | ✗ | ✗ |
| 20 | 2003 | 8 | 15 | 32.0 | 28.0 | ✗ | ✗ |
| 21 | 2002 | 7 | 43 | 20.4 | 17.0 | ✗ | ✗ |
| 22 | 2001 | 6 | 19 | 33.4 | 27.0 | ✗ | ✗ |
| 23 | 2002 | 6 | 35 | 33.2 | 30.0 | ✗ | ✗ |
| 24 | 2003 | 6 | 18 | 32.0 | 28.0 | ✗ | ✗ |
| 25 | 2002 | 5 | 12 | 20.4 | 17.0 | ✗ | ✗ |
| 26 | 2004 | 5 | 17 | 23.6 | 19.0 | ✗ | ✗ |
| 27 | 2000 | 4 | 15 | 22.1 | 17.0 | ✗ | ✗ |
| 28 | 2002 | 4 | 27 | 33.2 | 30.0 | ✗ | ✗ |
| 29 | 2002 | 4 | 14 | 18.3 | 16.0 | ✗ | ✗ |
| 30 | 2000 | 3 | 11 | 18.7 | 15.0 | ✗ | ✗ |
| 31 | 2003 | 3 | 20 | 16.2 | 14.0 | ✗ | ✗ |
| 32 | 2005 | 3 | 9 | 18.2 | 16.0 | ✗ | ✗ |
| 33 | 2005 | 2 | 10 | 33.3 | 31.0 | ✗ | ✗ |
| 34 | 2005 | 2 | 34 | 25.4 | 21.0 | ✗ | ✗ |
| 35 | 2003 | 1 | 5 | 16.2 | 14.0 | ✗ | ✗ |
| 36 | 2004 | 1 | 10 | 23.5 | 20.0 | ✗ | ✗ |
| 37 | 2005 | 1 | 5 | 18.2 | 16.0 | ✗ | ✗ |
| 38 | 2005 | 1 | 5 | 25.4 | 21.0 | ✗ | ✗ |
| 39 | 2001 | 0 | 10 | 20.3 | 15.0 | ✗ | ✗ |
| 40 | 2004 | 0 | 19 | 29.0 | 25.5 | ✗ | ✗ |
| 41 | 2004 | 0 | 24 | 23.6 | 19.0 | ✗ | ✗ |
| ***P*=42** | 2005 | 0 | 25 | 25.4 | 21.0 | ✗ | ✗ |
| Total | - | ***C*=416** | 747 | - | - | | ***h*-index=13** |
| Mean | - | 9.9 | 17.8 | 25.3 | 21.5 | | ***NSP*-index=10** |
| Median | - | 6.5 | 15.0 | 23.6 | 19.0 | | ***success*-index=7** |
| St.dev. | - | 11.0 | 10.1 | 6.8 | 7.1 | | |

**Tab. 2. Example of calculation of the *h*-, *NSP*- and the *success*-indices, considering the scientific production of two individual anonymous scientists (i.e. *S1*, a Chemist, and *S2*, a Computer Scientist) in the 2000-2005 time window. For the calculation of the *success*-index, $CT_i = \left(\tilde{r}_{JY}\right)_i$. In the two last columns of each table, ✓ and ✗ respectively denote papers included and not included in the *h*- or *success*-core. Articles were analyzed by means of the Scopus database, accessed on June 2011. For each group of publications, the total (if applicable), mean, median and standard deviation relating to $c_i$, $r_i$, $\left(\bar{r}_{JY}\right)_i$, and $\left(\tilde{r}_{JY}\right)_i$ values are reported at the bottom of the table.**

## 4. Operational properties of the *success*-index

In a previous work, we presented an overview of the most interesting operational properties of the *h*-index [Franceschini and Maisano, 2010a]. It is now proposed a discussion of the same properties regarding the *success*-index, highlighting its differences and (dis)advantages with respect to *h*. The parallelism between these two indicators seems appropriate for two main reasons: (1) they are almost comparable in terms of simplicity and immediateness of meaning, and (2) the *success*-index and *h* have the same dimension (*h* and the *success*-index $\in N_0$, i.e the set of natural numbers) since they both identify a sub-set of publications.

Tab. 3 proposes a list of the main simplifying assumptions behind the definition and use of the *success*-index. The fact that these assumptions may not actually be satisfied – e.g., in some cases self-citations or co-authorship may influence the *success*-index significantly – exposes the indicator to some criticism. Kosmulski (2011) suggested some refinements to reduce the previous limitations (for example, using citation and publication windows giving low importance to old citations or articles and vice-versa) but we think that they undermine the original simplicity and immediacy of *NSP* (and consequently the *success*-index's). It is worth noting that the previous assumptions apply to *h* as well. An additional limitation of *h* is that it does not allow comparisons between scholars from different fields [Franceschini and Maisano, 2010a].

| | |
|---|---|
| 1 | The diffusion/impact of an article is evaluated using the number of citations received. |
| 2 | Self citations do not increase the *success*-index significantly. |
| 3 | It is not essential to take the effect of multiple co-authorship into account. |
| 4 | Citations have the same importance, no matter what their age is or what the paper age is. |
| 5 | It is not necessary to distinguish between publications of different type (e.g., review, conference, journal articles). |
| 6 | The information used to determine the *success*-index – provided by public databases like WoS, GS or Scopus – is considered reliable, with no significant errors. In actual fact, using WoS, GS or Scopus can lead to different results since they use different sets of source journals and have different database limitations [Harzing and van der Wal, 2008]. |

**Tab. 3. Assumptions and simplifying hypotheses behind the definition and use of the *success*-index. The same assumptions apply to *h* as well.**

Regarding *h* and the *success*-index, the major operational properties are summarized in Tab. 4. They are individually analysed and discussed in the following subsections. For simplicity of exposition, we assume that each $CT_i$ is calculated as $\left(\tilde{r}_{JY}\right)_i$ (see Sect. 3).

| Property | Description |
|---|---|
| Simplicity of calculation | The calculation of *h* is based on the number of citations received by each of the publications at hand. The *success*-index also requires the knowledge of the citations made (or received) by a sample of "external" publications, for calculating the $CT_i$s. |
| Papers' passage in the core publications | The core papers are not fixed. Each paper can enter or go out of it depending on the citations earned over time. |
| No decreasing | A scientist's *h*-index will never decrease. Regarding the *success*-index, it depends on the way the comparison term is calculated. |
| Limited max value | The maximum value of the two indicators is limited by the number of papers. |
| Non compensation | The *h*- and *success*-indices do not follow the property of compensation, because they are not based on additive or multiplicative models for aggregating the publications and relevant citations. |
| Non (strict) monotony | The increase/decrease in number of citations of a paper is not necessarily associated to a corresponding increase/decrease of *h* (aggregated indicator), which is weakly monotonic. Regarding the *success*-index, it is weakly monotonic or non-monotonic depending on the way the comparison term is calculated. |
| Scale properties | *h* is defined on an ordinal scale, while the *success*-index is defined on a ratio scale. |
| Independence | An indicator IND is said to be independent if the following holds: If IND(*S1*) ≥ IND(*S2*), and one adds the same publication (with the same number of citations) to the publication lists of *S1* and *S2*, then IND(*S1'*) ≥ IND(*S2'*), where *S1'* and *S2'* denote scientists *S1* and *S2* to which this same publication has been added. While the *success*-index is independent, the *h*-index is not. |

**Tab. 4. Summary of the operational properties of the *h*- and *success*-index.**

### *4.1 Simplicity of calculation*

The *h*-index is very easy to determine since it only requires the knowledge of the number of citations received by the publications at hand. The *success*-index's calculation is more complex because it requires the knowledge of additional data (in this case $\left(\tilde{r}_{JY}\right)_i$).

### *4.2 Papers' passage in the core publications*

We define "core publications" a sub-set of the most relevant ones (accordingly to a certain bibliometric criterion). The *h*-core is given by the *h* most cited publications [Hirsch, 2005; Rousseau, 2006], while the *success*-core is given by the publications with a number of citations exceeding the corresponding comparison terms.

As regards *h*, a new publication entering the *h*-core may cause another publication to go out of it [Franceschini and Maisano, 2010a]. Regarding the *success*-index, each *i*-th publication of the *success*-core cannot go out of it, when the comparison term is fixed (this is the case of $\left(\tilde{r}_{JY}\right)_i$, see Fig. 4).

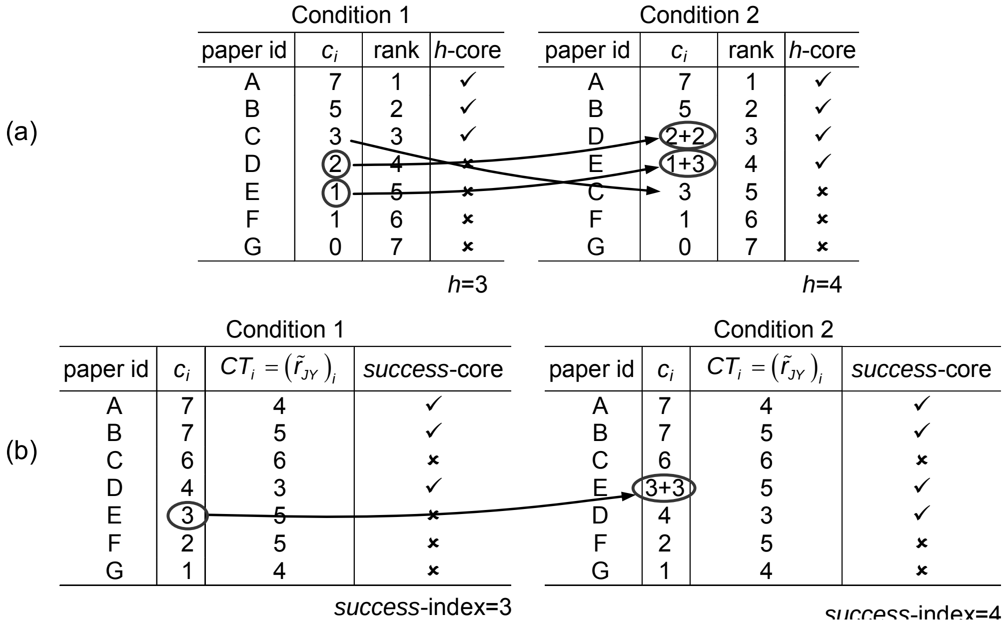


**Fig. 4. Example of papers' passage in the *h*- and the *success*-core. In the last column of each table, ✓ and × respectively denote papers included and not included in the *h*- or *success*-core. (a) Paper D and E increase their citation number, entering the *h*-core and making *h* increase from 3 to 4. At the same time, article C goes out of the *h*-core. (b) Paper E increases its citations, entering the *success*-core. In this case the comparison term of the *success*-index is calculated using $\left(\tilde{r}_{JY}\right)_i$. Since it does not change over time, any papers cannot go out of the *success*-core.**

### *4.3 No decreasing*

*h* cannot decrease over time. Also the *success*-index cannot decrease (when $CT_i$ is fixed), since no publications may come out from the *success*-core.

### *4.4 Limited max value*

*h* and the *success*-index are limited by the total number of publications, among which they identify a subset of core-publications, depending on some specific citation statistics. Obviously, both *h* and the *success*-index are included between 0 and *P* (*success*- and *h*-index $\in [0, P]$).

### *4.5 Non compensation*

The property of compensation can be studied when a system is represented by sub-indicators (in the specific case, publications and/or citations). If local changes (i.e., increase/decrease in performance) of sub-indicators may compensate each other – without making the aggregated indicator value change – then the aggregated indicator fulfils the property of compensation. This is typical of additive and multiplicative aggregation models [Franceschini et al., 2007]. Let us consider, for example, the indicator *total citations* (*C*) of a group of *P* publications, given by:

$$C = \sum_{i=1}^{P} c_i \, , \tag{7}$$

being $c_i$ the citations received by the *i*-th publication ($i = 1, \ldots, P$).

Given a set of publications and relevant citations, the increase in citations received by a publication can be compensated by the decrease of citations received by one or more other publications, so that

*C* does not change.

Neither *h*, nor the *success*-index follow the property of compensation, as – for example – these indicators are insensitive to the accumulation of new citations by the core-publications, or to the decrease in citations by the "off-core" publications (see Fig. 5).

Condition 1

| paper id | $c_i$ | rank | *h*-core | $CT_i = (\tilde{r}_{JY})_i$ | *success*-core |
|---|---|---|---|---|---|
| A | 9 | 1 | ✓ | 8 | ✓ |
| B | 8 | 2 | ✓ | 5 | ✓ |
| C | 5 | 3 | ✓ | 3 | ✓ |
| D | 3 | 4 | × | 1 | ✓ |
| E | 2 | 5 | × | 4 | × |
| F | 1 | 6 | × | 5 | × |

*h*=3 *success*-index=4

Condition 2

| paper id | $c_i$ | rank | *h*-core | $CT_i = (\tilde{r}_{JY})_i$ | *success*-core |
|---|---|---|---|---|---|
| A | 8 (-1) | 1 | ✓ | 8 | × |
| B | 8 | 2 | ✓ | 5 | ✓ |
| C | 5 | 3 | ✓ | 3 | ✓ |
| D | 4 (+1) | 4 | ✓ | 1 | ✓ |
| E | 2 | 5 | × | 4 | × |
| F | 1 | 6 | × | 5 | × |

*h*=4 *success*-index=3

**Fig. 5. *h* and the *success*-index do not fulfil the compensation property. ✓ and × respectively denote papers included and not included in the *h*- or *success*-core. Two conditions are exemplified: in condition 2, paper A has one citation less and paper D has one citation more than in condition 1; nevertheless, in condition 2, *h* and the *success*-index are respectively larger and lower than in condition 1. According to *h* and the *success*-index, compensation is not fulfilled. In condition 2, the higher citation rate of paper D is not counterbalanced by the lower citation rate of paper A.**

### *4.6 Non (strict) monotony*

If a system is represented by different sub-indicators (publications and/or relevant citations), aggregated into a synthetic indicator, and if the increase/decrease of one sub-indicator is not associated to the increase/decrease of the aggregated indicator, then the last one does not fulfil the condition of (strict) monotony (Franceschini et al., 2007). The *h*- and the *success*-indices are only weakly monotonic, since the introduction of a new citation or publication is a necessary but not sufficient condition to raise them (see Fig. 6).

Condition 1

| paper id | $c_i$ | rank | *h*-core | $CT_i = (\tilde{r}_{JY})_i$ | *success*-core |
|---|---|---|---|---|---|
| A | 10 | 1 | ✓ | 5 | ✓ |
| B | 9 | 2 | ✓ | 5 | ✓ |
| C | 4 | 3 | ✓ | 5 | × |
| D | 3 | 4 | × | 5 | × |

*h*=3 *success*-index=2

Condition 2

| paper id | $c_i$ | rank | *h*-core | $CT_i = (\tilde{r}_{JY})_i$ | *success*-core |
|---|---|---|---|---|---|
| A | 10 | 1 | ✓ | 5 | ✓ |
| B | 9 | 2 | ✓ | 5 | ✓ |
| C | 5 (+1) | 3 | ✓ | 5 | × |
| D | 3 | 4 | × | 5 | × |

*h*=3 *success*-index=2

**Fig. 6. *h* and the *success*-index do not fulfil the property of strict monotony. The increase in the number of citations of paper C (from 4 to 5) is not associated to a corresponding increase of *h* and the *success*-index. Precisely, *h* and the *success*-index are weakly monotonic. ✓ and × respectively denote papers included and not included in the *h*- or *success*-core.**

### *4.7 Scale properties*

The *h*-index scale has only the *equivalence* and *ordinal* properties [Franceschini and Maisano, 2010a]. This means that *h* makes it possible to compare and order groups of publications (typically those associated to a scientist), but tells us nothing about their "distance". For example, it can be stated that a scientist with *h*=10 is likely to be better than another with *h*=2, however none can say that the "gap" between two scientists with *h*-indices 16 and 18 is equivalent to the gap between two scientists with *h*-indices 1 and 3. Actually, Hirsch empirically showed that, considering a scholar's scientific production, the total number of citations ($C$) is approximately proportional to $h^2$ [Hirsch, 2005]:

$$C \approx ah^2 \tag{8}$$

Thus, $h$ value is roughly proportional to $C^{1/2}$. The same result is confirmed by the study of Egghe and Rousseau (2006): they prove that $h$ is proportional to $C^{1/\alpha}$, where α equals the exponent in the law of Lotka, which most classical value is 2. According to Eq. 8, Fig. 7 illustrates the values of $h$ depending on the value of $C$. For the purpose of simplicity $a$ is assumed to be unitary.

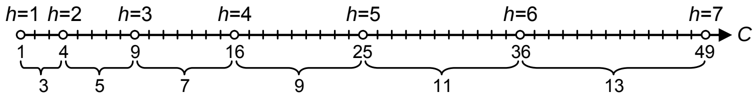


**Fig. 7. *h* values represented on the *C* axis. For the purpose of simplicity we assumed *a*=1, so $h \approx C^{1/2}$. The distance (in terms of citations) between two consecutive *h*-classes is not constant [Franceschini and Maisano, 2010a].**

The scale of the *success*-index is much richer because it has the *ratio* property [Roberts, 1979]. It is worth noting that the *absolute zero* of the *success*-index's scale corresponds to the condition in which there are no successful publications. If a scientist has *success*-index=4 it can be said that his/her performance, according to this indicator, is twice that one of a scientist with *success*-index=2.

Additionally, it is well known that $h$ does not necessarily reflect compositions of the input publication portfolios (with the corresponding citations). In other terms, the union of two groups of publications ($a$ and $b$), with $h$-indexes of 3 and 5 respectively, does not necessarily originate a third group ($a + b$) with $h$-index of 3 + 5 = 8 [Egghe, 2008]. On the other hand, the *success*-index makes this type of composition possible, thanks to the ratio property of its scale. For this reason, the union of two groups of publications with *success*-index of 2 and 5 (with no common publications) will always originate a third group of publications with *success*-index of 2 + 5 = 7 (see Fig. 8). This simple property is very useful for extending the use of the *success*-index to multi-disciplinary research groups and institutions. Of course, the (reasonable) assumption behind this reasoning is that the intersection between the groups of publications (from different disciplines) to be joined is null.

String $a$

| paper id | $c_i$ | rank | $h$-core | $CT_i = (\tilde{r}_{JY})_i$ | *success*-core |
|---|---|---|---|---|---|
| A | 7 | 1 | ✓ | 3 | ✓ |
| B | 5 | 2 | ✓ | 4 | ✓ |
| C | 5 | 3 | ✓ | 5 | × |
| D | 3 | 4 | × | 4 | × |
| E | 2 | 5 | × | 6 | × |
| F | 0 | 6 | × | 4 | × |

$h$=3 *success*-index=2

String $b$

| paper id | $c_i$ | rank | $h$-core | $CT_i = (\tilde{r}_{JY})_i$ | *success*-core |
|---|---|---|---|---|---|
| G | 18 | 1 | ✓ | 5 | ✓ |
| H | 16 | 2 | ✓ | 6 | ✓ |
| I | 12 | 3 | ✓ | 4 | ✓ |
| J | 9 | 4 | ✓ | 5 | ✓ |
| K | 5 | 5 | ✓ | 6 | × |
| L | 5 | 6 | × | 4 | ✓ |
| M | 3 | 7 | × | 3 | × |
| N | 2 | 8 | × | 6 | × |
| O | 1 | 9 | × | 5 | × |
| P | 1 | 10 | × | 4 | × |

$h$=5 *success*-index=5

String $a + b$

| paper id | $c_i$ | rank | $h$-core | $CT_i = (\tilde{r}_{JY})_i$ | *success*-core |
|---|---|---|---|---|---|
| G | 18 | 1 | ✓ | 5 | ✓ |
| H | 16 | 2 | ✓ | 6 | ✓ |
| I | 12 | 3 | ✓ | 4 | ✓ |
| J | 9 | 4 | ✓ | 5 | ✓ |
| A | 7 | 5 | ✓ | 3 | ✓ |
| K | 5 | 6 | × | 6 | × |
| L | 5 | 7 | × | 4 | ✓ |
| C | 5 | 8 | × | 5 | × |
| B | 5 | 9 | × | 4 | ✓ |
| M | 3 | 10 | × | 3 | × |
| D | 3 | 11 | × | 4 | × |
| E | 2 | 12 | × | 6 | × |
| N | 2 | 13 | × | 6 | × |
| O | 1 | 14 | × | 6 | × |
| P | 1 | 15 | × | 5 | × |
| F | 0 | 16 | × | 4 | × |

$h$=3 *success*-index =7

**Fig. 8. *h* does not reflect compositions of the input citation strings. String *a* (*h*=3) is joined to string *b* (*h*=5), obtaining a string with corresponding *h*=5. On the other hand, there is direct relationship between the *success*-index values of the single input citation string, and the *success*-index values of the corresponding joined strings. ✓ and × respectively denote papers included and not included in the *h*- or *success*-core.**

### *4.8 Independence*

As a supplement to the reasoning on the composition of publication-citations lists, it is appropriate to introduce the property of independence. An indicator IND is said to be independent if the following holds: Scientist *S1* (represented by his/her publication-citation list) is considered to be at least as good as scientist *S2* (also represented by a publication-citation list), hence IND(*S1*) ≥ IND(*S2*), and one adds the same publication(s) (with the same number of citations) to the publication lists of *S1* and *S2*, then IND(*S1'*) ≥ IND(*S2'*), where *S1'* and *S2'* denote scientists *S1* and *S2* to which the same publication(s) have been added [Rousseau and Ye, 2011].

While the *success*-index is independent, the *h*-index is not (see the exemplification in Fig. 9).

### *4.9 Further remarks*

It is important to point out that in the previous discussion about the *success*-index properties, it was assumed $CT_i = (\tilde{r}_{JY})_i$. It is worth remarking that for different $CT_i$ constructions, some of the above properties may change. For example, if $CT_i$ is calculated on the basis of the citations received (not made) by a reference sample of publications, then it will be subject to a possible increase over time, and so on.

Scientist *S1*

| paper id | $c_i$ | rank | *h*-core | $CT_i = (\tilde{r}_{JY})_i$ | *success*-core |
|---|---|---|---|---|---|
| A | 6 | 1 | ✓ | 4 | ✓ |
| B | 4 | 2 | ✓ | 3 | ✓ |
| C | 4 | 3 | ✓ | 5 | × |
| D | 4 | 4 | ✓ | 2 | ✓ |
| E | 1 | 5 | × | 3 | ✓ |
| F | 1 | 6 | × | 2 | × |
| | | | *h*=4 | | *success*-index=4 |

Additional papers

| paper id | $c_i$ | rank | *h*-core | $CT_i = (\tilde{r}_{JY})_i$ | *success*-core |
|---|---|---|---|---|---|
| W | 5 | - | - | 6 | × |
| Z | 5 | - | - | 3 | ✓ |

Scientist *S2*

| paper id | $c_i$ | rank | *h*-core | $CT_i = (\tilde{r}_{JY})_i$ | *success*-core |
|---|---|---|---|---|---|
| G | 7 | 1 | ✓ | 4 | ✓ |
| H | 5 | 2 | ✓ | 5 | × |
| I | 5 | 3 | ✓ | 2 | ✓ |
| J | 2 | 4 | × | 3 | × |
| K | 1 | 5 | × | 4 | ✓ |
| | | | *h*=3 | | *success*-index=3 |

Scientist *S1'*

| paper id | $c_i$ | rank | *h*-core | $CT_i = (\tilde{r}_{JY})_i$ | *success*-core |
|---|---|---|---|---|---|
| A | 6 | 1 | ✓ | 4 | ✓ |
| W | 5 | 2 | ✓ | 6 | × |
| Z | 5 | 3 | ✓ | 3 | ✓ |
| B | 4 | 4 | ✓ | 3 | ✓ |
| C | 4 | 5 | × | 5 | × |
| D | 4 | 6 | × | 2 | ✓ |
| E | 1 | 7 | × | 3 | ✓ |
| F | 1 | 8 | × | 2 | × |
| | | | *h*=4 | | *success*-index=5 |

Scientist *S2'*

| paper id | $c_i$ | rank | *h*-core | $CT_i = (\tilde{r}_{JY})_i$ | *success*-core |
|---|---|---|---|---|---|
| G | 7 | 1 | ✓ | 4 | ✓ |
| H | 5 | 2 | ✓ | 5 | × |
| W | 5 | 3 | ✓ | 6 | × |
| Z | 5 | 4 | ✓ | 3 | ✓ |
| I | 5 | 5 | ✓ | 2 | ✓ |
| J | 2 | 6 | × | 3 | × |
| K | 1 | 7 | × | 4 | ✓ |
| | | | *h*=5 | | *success*-index=4 |

**Fig. 9. *h* does not fulfil the property of independence, while the *success*-index does. The addition of the same two papers (i.e., W and Z) to the publication-citation lists of two scientists (*S1* and *S2*) may produce a "rank reversal" in the values of the *h*-index, not in the *success*-index's. In detail, *h*(*S1*)>*h*(*S2*) and *h*(*S1'*)<*h*(*S2'*), while *success*(*S1*)>*success*(*S2*) and *success*(*S1'*)>*success*(*S2'*).**

## 5. Conclusions and discussion

This paper focused on a generalized index based on the Kosmulski's *NSP*-index, which stands out for its particular simplicity and immediacy of meaning, comparable to those of the *h*-index. One of the most interesting features of *NSP* is that it enables the citation impact normalization of the differences among scientific fields at the level of individual publication. This feature makes *NSP* suitable for the evaluation of groups of scientific publications that are not necessarily within the same discipline. For example, the publications of eclectic scientists who are involved in more than one discipline (e.g., most of the bibliometrists), or – more realistically – publications associated with multidisciplinary research institutions or organizations.

The *NSP*-index, as well as many different normalized impact indicators, is based on the comparison between the citations received from a publication and a comparison term, which is – in the original formulation of Kosmulski – given by the publication's number of bibliographic references. While being very simple, this formulation has two main drawbacks: (1) for a certain publication, the estimation of the propensity to cite is ineffective in terms of statistical significance, and (2) *NSP* is prone to manipulation. We therefore suggested to modify *NSP* (into the *success*-index) by replacing the comparison term with a more appropriate indicator of propensity to cite, which has to be determined on the basis of a reference sample of publications. There are many options for defining the indicator of propensity to cite and selecting the publication sample. For example, the indicator can be given by the mean/median citations received (or made) and the sample can be given by the "neighbours" (i.e., the article cited) of the publication of interest. This variant makes the calculation of the *success*-index much more complex, but it does not undermine the indicator's immediacy of meaning.

The second part of the paper analyzed the properties of the *success*-index, highlighting the similarities and differences with *h*. We point out that the scale of measurement of the *success*-index is much richer (ratio scale) than the *h*-index's (ordinal scale). This peculiarity makes the indicator *success*-index much more versatile than *h*, because of its adaptability to different contexts, without requiring special adjustments. To clarify this concept, here follows a list of possible uses of the *success*-index, which cannot be easily achieved by *h*:

1. Individual scientists of different seniority (including those involved in more than one scientific field) can be compared by means of the number of "successful papers" per career year (i.e., *success*-index/*Y*, being *Y* the total career years);
2. A simple indicator for the evaluation and comparison of scientific journals (even considering different scientific fields) can be the percentage of "successful papers" on the total of published articles (i.e., *success*-index/*P*, being *P* the number of articles published on a journal in a specific time period and the *success*-index the corresponding "successful papers");
3. Some simple synthetic indicators for evaluating and comparing entire research institutions/organizations – independently on the disciplines involved, staff number, amount of funding received, etc. – can be (i) the average number of successful papers per head, (ii) the average number of successful papers per unit of investment, (iii) the percentage of successful publications, etc. [Leydesdorff and Shin, 2011].

These and other possible applications of the *success*-index denote the great potential of this indicator and deserve the attention of future research.

## References

Amin, M. Mabe, M. (2000). Impact Factors: Use and Abuse. Elsevier Science, Perspectives in Publishing, n. 1

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

Bornmann, L. (2011). Mimicry in science?. Scientometrics, 86(1), 173–177.

Bornmann, L., de Moya Anegon, F. (2011). Some interesting insights from aggregated data published in the World Report SIR 2010. Journal of Informetrics, 5(3), 486–488.

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

Davis, P. (2010). Reference List Length and Citations: A Spurious Relationship. Retrieved on August 1, 2011, from http://scholarlykitchen.sspnet.org/2010/08/18/reference-list-length-and-citations-a-spurious-relationship/.

Egghe, L., Rousseau, R. (1990). Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science. Elsevier Science Publishers. Retrieved on November 21, 2011, from http://hdl.handle.net/1942/587.

Egghe, L., Rousseau, R. (2006). An informetric model for the Hirsch-index, Scientometrics, 69(1), 121–129.

Egghe, L. (2008). The influence of merging on h-type indices. Journal of Informetrics, 2(3), 252–262.
Franceschini, F., Galetto, M., Maisano, D. (2007). Management by Measurement: Designing Key Indicators and Performance Measurements. Springer: Berlin.
Franceschini, F., Maisano, D. (2010a). Analysis of the Hirsch index's operational properties. European Journal of Operational Research, 203(2), 494–504.
Franceschini F, Maisano D. (2010b). The Hirsch spectrum: a novel tool for analysing scientific journals. Journal of Informetrics. 4(1): 64–73.
Franceschini, F., Maisano, D. (2010c). The citation triad: An overview of a scientist's publication output based on Ferrers diagrams. Journal of Informetrics, 4(4), 503–511.
Franceschini, F., Maisano, D. (2011) Bibliometric positioning of scientific Manufacturing journals: A comparative analysis. Scientometrics, 86(2): 463–485.
Garfield, E. (1979a). Citation indexing. Its theory and application in science, technology and humanities. New York: Wiley
Garfield, E. (1979b). Is citation analysis a legitimate evaluation tool? Scientometrics, 1(4), 359–375.
Garfield E. (2005). Agony and the ecstasy – the history and meaning of the impact factor. Proceedings of the International Congress on Peer review and Biomedical Publication. Chicago, USA, September 16, 2005.
Glänzel, W., Schubert, A., Thijs, B., Debackere, K. (2011). A priori vs. a posteriori normalisation of citation indicators. The case of journal ranking. Scientometrics, 87(2), 415–424.
Guns, R., Rousseau. R. (2010). New journal impact indicators take references into account: a comparison. ISSI Newsletter, 6(1), 9–14.
Harzing, A.W., van der Wal, R. (2008). Google Scholar as a new source for citation analysis. Ethics in Science and Environmental Politics, 8(11), 61–73.
Haslam, N., Ban, L., Kaufmann, L., Loughnan, S., Peters, K., Whelan, J., Wilson, S. (2008). What makes an article influential? Predicting impact in social and personality psychology. Scientometrics, 76(1), 169–185.
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, 102, 16569–16572.
Kochen, M. (1974). Principles of Information Retrieval, Los Angeles: Melville, p.21.
Kosmulski, M. (2009). New seniority-independent Hirsch-type index. Journal of Informetrics, 3(4), 341–347.
Kosmulski, M. (2011). Successful papers: A new idea in evaluation of scientific output. Journal of Informetrics, 5(3), 481–485
Krampen, G. (2010). Acceleration of citing behavior after the millennium? Exemplary bibliometric reference analyses for psychology journals. Scientometrics, 83(2), 507–513.
Jackson, M.O., Rogers B.W. (2007). Meeting strangers and friends of friends: How random are social networks? American Economic Review, 97(3), 890–915.
JCQAR – Joint Committee on Quantitative Assessment of Research (2010). Citation Statistics. Retrieved on August 1, 2011, from http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf.
Iglesias, J.E, Pecharromán, C. (2007). Scaling the h-index for different scientific ISI fields. Scientometrics, 73(3), 303–320.
Larivière, V., Archambault, É., Gingras, Y., Wallace, M.L. (2008). The fall of uncitedness, Book of Abstracts of the 10th International Conference on Science and Technology Indicators (ISSI), pp. 279-282.
Leydesdorff, L., Bornmann, L. (2011). Integrated impact indicators compared with impact factors: An alternative research design with policy implications. Journal of the American Society for Information Science and Technology, 62(11), 2133–2146.
Leydesdorff, L., Shin, J.C. (2011). How to Evaluate Universities in Terms of Their Relative Citation Impacts: Fractional Counting of Citations and the Normalization of Differences Among Disciplines. Journal of the American Society for Information Science and Technology, 62(6), 1146–1155.
Leydesdorff, L., Opthof, T. (2010). Normalization at the field level: fractional counting of citations. Journal of Informetrics, 4(4), 644–646.
Mingers, J. (2009). Measuring the research contribution of management academics using the Hirsch-index. Journal of the Operational Research Society, 60(9), 1143–1153.
Mitchell, D.W. (2004). More on spreads and non-arithmetic means. The Mathematical Gazette, 88(511), 142–144.
Moed, H.F. (2010a). Measuring contextual citation impact of scientific journals. Journal of Informetrics, 4(3), 265–277.
Moed, H.F. (2010b). CWTS crown indicator measures citation impact of a research group's publication

oeuvre. Journal of Informetrics, 3(3), 436–438.
Pinski, G., Narin., F. (1976). Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. Information Processing and Management, 12(5), 197–312.
Ravichandra Rao, I.K. (2011). Relations among the number of Citations, References and Authors: Revisited. Proceedings of COLLNET 2011, 7th International Conference on Webometrics, Informetrics and Scientometrics (WIS), 20-23 September, 2011, Istanbul, Turkey.
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. (2006). New developments related to the Hirsch index. Science Focus 1(4), 23–25.
Rousseau, R., Ye, F.Y. (2011). A simple impact measure and its evolution over time. Journal of Library & Information Studies, 9(2).
SCImago Research Group (2009). SCImago Institutions Rankings (SIR): 2009 world report. Retrieved on August 1, 2011, from http://www.scimagoir.com/pdf/sir_2009_world_report.pdf.
Small, H. (2004). On the shoulders of Robert Merton: Towards a normative theory of citation. Scientometrics, 60(1), 71–79.
Small, H., Sweeney, E. (1985). Clustering the Science Citation Index Using Co-Citations I. A Comparison of Methods. Scientometrics 7(3-6), 391–409.
Waltman, L., van Eck, N.J., Van Leeuwen, T.N., Visser, M.S., Van Raan, A.F.J. (2011a). Towards a new crown indicator: Some theoretical considerations. Journal of Informetrics, 5(1), 37–47.
Waltman, L., Yan, E., van Eck, N.J. (2011b). A recursive field-normalized bibliometric performance indicator: an application to the field of library and information science. Scientometrics, 89(1), 301–314.
Williamson, J.R. ( 2009). My h-index turns 40: My midlife crisis of impact, ACS Chemical Biology 4(5), 311-313.
Zhou, P., Leydesdorff, L. (2011). Fractional counting of citations in research evaluation: A cross- and interdisciplinary assessment of the Tsinghua University in Beijing. Journal of Informetrics, 5(3), 360–368.
Zitt, M. (2010). Citing-side normalization of journal impact: A robust variant of the Audience Factor. Journal of Informetrics, 4(3), 392–406.
Zitt, M., Small, H. (2008). Modifying the journal impact factor by fractional citation weighting: The audience factor. Journal of the American Society for Information Science and Technology, 59(11), 1856–1860.