Scientometrics (2013) 96:85–101
DOI 10.1007/s11192-012-0887-2

Evaluating research institutions: the potential of the *success*-index

Fiorenzo Franceschini · Domenico Maisano · Luca Mastrogiacomo

Received: 11 August 2012 / Published online: 30 October 2012

Abstract Similarly to the *h*-index and other indicators, the *success*-index is a recent indicator that makes it possible to identify, among a general group of papers, those of greater citation impact. This indicator implements a field-normalization at the level of single paper and can therefore be applied to multidisciplinary groups of articles. Also, it is very practical for normalizations aimed at achieving the so-called size-independency. Thanks to these (and other) properties, this indicator is particularly versatile when evaluating the publication output of entire research institutions. This paper exemplifies the potential of the *success*-index by means of several practical applications, respectively: (i) comparison of groups of researchers within the same scientific field, but affiliated with different universities, (ii) comparison of different departments of the same university, and (iii) comparison of entire research institutions. A sensitivity analysis will highlight the *success*-index's robustness. Empirical results suggest that the *success*-index may be conveniently extended to large-scale assessments, i.e., involving a large number of researchers and research institutions.

Keywords *Success*-index · Hirsch index · Field normalization · Citation propensity · Groups of researchers · Research institutions

Introduction

Research assessments often concern the scientific output of groups of researchers, such as those affiliated with departments, faculties, disciplinary sectors, universities or other research institutions. In several countries around the world, bibliometric evaluations of the research output of universities and research institutions are commonly adopted for the purpose of allocating public funds in a meritocratic way (Wang et al. 2012). According to

F. Franceschini · D. Maisano (✉) · L. Mastrogiacomo
DIGEP (Department of Management and Production Engineering),
Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy
e-mail: domenico.maisano@polito.it

F. Franceschini
e-mail: fiorenzo.franceschini@polito.it

the same logic, funds are often distributed among departments and groups of researchers within universities.

While research meritocracy is (almost) universally accepted, there are many disputes on the choice of the most appropriate tools. Typical assessments involve set of scientists from multiple research fields, hence the use of field-normalized indicators is essential. Otherwise, scientists from areas with a great citation propensity (e.g., Biology and Medicine) would be favoured to the detriment of other scientists from areas with an opposite attitude (e.g., Mathematics) (Garfield 1979). The use of field-normalized indicators is also necessary for evaluations of scientists within the same field, in the following cases: (1) scientists specialized in sub-disciplines with significant differences in terms of citation propensity (Glänzel et al. 2011), and (2) scientists involved in multiple (sub-)disciplines (e.g., mathematicians or computer scientists actively involved in bibliometrics).

Additional general requirements of indicators are simplicity and immediateness of meaning. In fact, regardless of their efficiency, indicators of complicate meaning often tend to be set aside (Franceschini et al. 2007).

In a recent article, Franceschini et al. (2012a) introduced the *success*-index of a generic group of scientific papers, defined as the number of papers with a number of citations greater than or equal to CT_i, i.e., a generic comparison term associated with a generic i-th paper. 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. Because of the fact that the (non-)success status of a specific paper is determined independently on the other papers of interest, the *success*-index can be applied to groups of papers from different disciplines.

Considering the *success*-index from a broader perspective, it can be seen that—given a generic set of publications—this indicator allows to select an "elite" subset. This selection can also be made by other indicators in the literature: e.g., let us consider the h-core approach (Burrell 2007), the selection by π-indicator (Vinkler 2011), the *characteristic scores and scales* (CSS) method (Glänzel 2011) or the ESI's Highly Cited Papers method (ISI Web of Knowledge 2012). We remark that, differently from the *success*-index, the aforementioned methods require that the set of publications examined are preliminarily divided into scientific (sub-)disciplines.

Thanks to some peculiar properties described in detail in (Franceschini et al. 2012a), the *success*-index is very versatile for other normalizations aimed at obtaining the so-called size-independency. Given a general group of (multidisciplinary) papers and the same capacity of producing "successful" papers, it is reasonable to assume that the *success*-index should increase proportionally with the different types of "resources" deployed. Different types of normalizations can be made depending on the type of "resource" considered, for example:

- when comparing groups of researchers with different staff number (N), a practical size-independent indicator is the *success*-index *per capita*, i.e., *success*/N;
- when comparing groups of researchers with different seniority (i.e., years of activity, Y), a possible indicator of annual performance is *success*/Y;
- when comparing journals with a different number of articles (P), a possible indicator of average impact per article is *success*/P;
- when comparing research institutions with different amount of funding, a possible indicator of efficiency in the scientific output is *success*/\$, being \$ the amount of funding received in a certain period.

In general, the *success*-index is suitable for evaluating the publication output of multidisciplinary research institutions. This potential is not assured by many other indicators. For example, let us consider the h-index (Hirsch 2005):

- h is not field-normalized;
- normalizations aimed at obtaining size-independency are complicated by the fact that h does not tend to be directly proportional to the "resources" invested;
- h is not appropriate for compositions of publication portfolios.

For more information on the points above, please refer to Courtault and Hayek (2008), Franceschini and Maisano (2010), Franceschini and Maisano (2011) and Franceschini et al. (2012b).

The aim of this paper is to show how effective the *success*-index is when evaluating groups of researchers and research institutions. Three different types of practical applications will be presented:

1. analysis of groups of researchers within the same disciplinary sector, but affiliated with different universities;
2. analysis of scientists from different departments of the same university;
3. analysis of different universities, based on their total publication output.

In all the three cases, groups of scientists have different staff number. Analysis is limited to papers issued in the period from 2004 to 2008. In order to avoid distortions in terms of different maturation time of citation impact, the comparison will be performed by defining a fixed citation time-window for each paper.

The remaining of the paper is organized as follows. "Brief remarks on the *success*-index" recalls the procedure for constructing the *success*-index, with particular attention to the problem of determining the CT_i values. "Data collection and methodology" presents the analysis methodology, defining the indicators in use and the data collection procedure. "Results" illustrates and comments analysis results. "Sensitivity analysis" shows some guidelines for checking the robustness of *success*-index by a sensitivity analysis. Finally, the conclusions are given, summarising the original contribution of the paper.

Brief remarks on the *success*-index

According to the formal definition of the *success*-index (Franceschini et al. 2012a), a score is associated to each (i-th) of the (P) publications of interest:

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

where c_i are the citations obtained and CT_i is the comparison term of the i-th publication. The *success*-index is given by:

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

The most complicated operation when constructing the *success*-index is determining the CT_i value of each (i-th) paper, i.e., the estimator of the "citation potential". To do this, there are different possible approaches, which are borrowed from the existing literature on field-normalized indicators. Despite their differences, most of them are based on 3 steps:

1. Selecting a reference sample of papers, which are homologous to the one of interest, to be used for estimating its citation propensity. The sample must be large enough to be statistically representative but, at the same time, should not be "polluted by outsider

articles", such as papers from other (sub-)disciplines. This is probably the most difficult operation, since it represents a still open problem in bibliometrics. Some approaches are "classification dependent", as they are based on superimposed classifications, such as ISI subject categories (Waltman et al. 2011); others are "adaptive", since the sample is determined considering the "neighbourhood" of the publication(s) of interest—generally consisting of the set of publications that cite or are cited by them (Zitt and Small 2008; Leydesdorff and Opthof 2010; Moed 2010a; Zitt 2010; Glänzel et al. 2011).

2. Estimating the citation propensity by using the distribution of (i) the number of references given or (ii) the citations obtained by the publications of the reference sample. We remark that the condition that the average number of references given should be equal to the average number of citations obtained is valid only within relatively closed information fields (Vinkler 2004; Franceschini et al. 2012a).
3. Defining a suitable (central tendency) indicator for obtaining CT_i from the distribution of interest, e.g., mean, median, harmonic mean, percentiles, etc. We remark that, being based on the comparison between the citations obtained by individual articles and some citation thresholds determined by the mean (or another central tendency indicator) of many papers, the *success*-index can somewhat be statistically questionable.

We remark that the identification of the most appropriate estimators of citation propensity is a general open issue when dealing with field-normalized indicators. The main obstacle is probably the lack of an absolute validation procedure (Radicchi and Castellano 2012). It should also be noted that each normalization procedure requires a time-window for counting the citations obtained by the paper(s) of interest, which ought to be consistent with the time-window used to determine the estimator(s) of citation propensity (Moed 2010a). This aspect will be illustrated in more detail in "The problem of calculating CT_i".

Data collection and methodology

Thanks to the aforementioned characteristics, the *success*-index allows comparisons between research groups, regardless of their discipline(s) of interest and staff number (*N*). In our study, three different types of bibliometric comparisons will be made (see Table 1):

1. among four groups of researchers in the same *Disciplinary Sector*[1] (SSD, according to the Italian acronym) of *Production Technologies and Manufacturing Systems*, affiliated with four different Italian universities (i.e., *Politecnico di Milano*, *Politecnico di Torino*, *Università di Napoli Federico II* and *Università di Palermo*);
2. between the tenured staff of two departments (i.e., Depts. of Chemistry and Mathematics) of *Politecnico di Torino*;
3. among the tenured staff of three Italian technical universities (i.e., *Politecnico di Bari*, *Politecnico di Milano* and *Politecnico di Torino*).

The three types of analysis are limited to papers issued within the five-year period from 2004 to 2008. In all three cases, the *N* of the groups of researchers is variable and their members were identified through public directories (MIUR 2012). The Scopus database was queried on July 2012, retrieving (1) the total scientific output of each group (given by the union of the contributions of individual researchers) and (2) the corresponding citation statistics (Scopus 2012).

[1] In Italy each university scientist belongs to one specific disciplinary sector, 370 in all. Complete list accessible at http://cercauniversita.cineca.it/php5/settori/index.php, last accessed on July, 2012.

Table 1 Summary of the bibliometric evaluations to be performed

Analysis type	Groups compared		N	Period of interest	Indicators in use
1) Comparison among groups of researchers from the same SSD (i.e., *Production Technologies and Manufacturing Systems*)	G1	from *Politecnico di Milano*	20	2004–2008	*P*, *C*, *CPP*, *h*-index and *success*-index
	G2	from *Politecnico di Torino*	15		
	G3	from *Università di Napoli Federico II*	12		
	G4	from *Università di Palermo*	13		
2) Comparison between the scientific output of two departments	D1	Dept. of Mathematics	60	Ibid.	Ibid.
	D2	Dept. of Chemistry	88		
	(both from *Politecnico di Torino*)				
3) Comparison among entire research universities, on the basis of the whole publication output	U1	*Politecnico di Torino*	837	Ibid.	Ibid.
	U2	*Politecnico di Milano*	1,388		
	U3	*Politecnico di Bari*	316		

For each group of researchers are reported the abbreviation used hereafter in the text and the staff number (*N*)

For each group, we conventionally refer to the number of researchers observed at the moment of the analysis (i.e., July 2012), considering it as a constant throughout the whole period of interest. However, we are aware that the *N* of the single groups of researchers may be subject to small variations over time—e.g., due to staff retirements, new hires, transfers, etc.

The documents analysed consist of articles from journals indexed by Scopus and do not include other sources—such as conference proceedings, books, book chapters, or some minor or non–English journals—not covered by the database.

The risk of analysis distortions because of homonymies or other ambiguities is minimized by the database query method, which includes researchers' names and affiliations.

The comparison among groups of researchers will be focussed on the calculation of the *success*-index relating to the scientific output (1) at annual level and (2) during the whole period of interest. In addition to the *success*-index, other traditional non-field-normalized indicators will be constructed, i.e., *P* (total number of papers), *C* (total number of citations), *CPP* (average citations per paper) and the *h*-index relating to the totality of the paper of each group of researchers.

To enable comparisons between indicators concerning papers of different age (i.e., different maturation time of citation impact), citations are counted using a fixed three-year time-window. For instance, citations obtained by papers published in 2004 are counted starting from 1 year after 2004, i.e., 2005, 2006 and 2007.

The problem of calculating CT_i

As regards the construction of the *success*-index, special attention deserves the choice of a suitable procedure for calculating the CT_i values relating to the papers examined. In our opinion, a sound procedure—hereafter referred to as "wished procedure"—would be structured as follows (see Fig. 1):

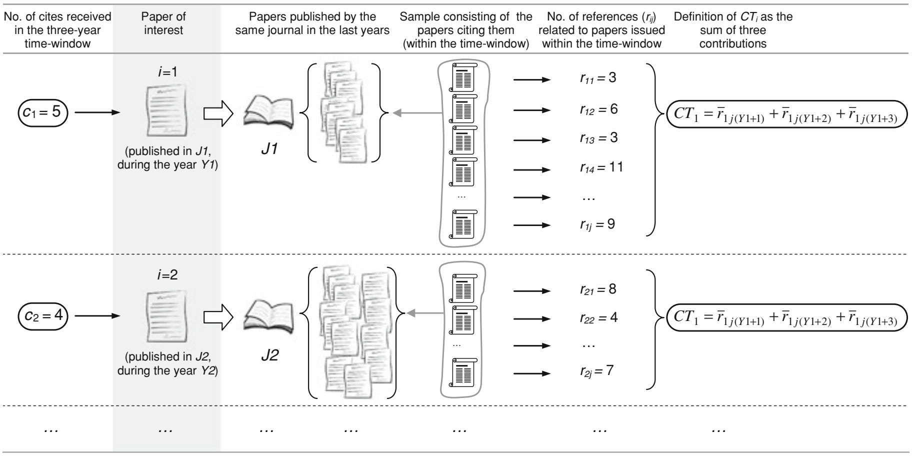

Fig. 1 Schematic representation of the "wished procedure" for constructing the CT_i values associated with the papers of interest (highlighted in *gray*). The data circled—i.e., number of citations obtained (c_i) and CT_i—can be used to determine the (non-)success status of each i-th paper

1. Identify a sample of "homologous" papers (from the whole literature), by selecting those papers that cite—during the 3-year window for citation count—other articles published by the same journal of the paper of interest, in the 10 years before the time-window. This conventional ten-year period was chosen in order to obtain a sample of adequate size. Also, this selection is based on the reasonable assumption that articles citing other articles from the same journal are relatively similar in terms of average propensity to cite. We observe that this assumption does not hold in the case of multidisciplinary journals (e.g., Nature and Science) and other journals with a broad scope (e.g., New England Journal of Medicine and Physical Review Letters).
2. Determine the number of references given by the articles of the sample (i.e., r_{ij} values, being j a generic article of the sample related to the i-th paper of interest). We remark that indicators based on the distribution of the references given—rather than the citations obtained—by the articles of the reference sample have several advantages (Franceschini et al. 2012a, c):

   – the number of references given is fixed over time, while the number of citations obtained tends to increase and requires a certain period to stabilize;
   – 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 obtained. The estimation will therefore be less subject to fluctuations, especially in the case of relatively small samples of publications (Albarrán and Ruiz-Castillo 2011).

We focus the attention on the fact that the number of citations obtained by each i-th paper is limited to (1) citations from documents issued in the (3-year) time-window related to the paper of interest and (2) citations from documents indexed by the database in use (in our case, Scopus). Citations from "external" sources—e.g., monographs, dissertations, articles on minor scientific journals that are not indexed by the database—are not taken into account. Thus, for a consistent estimation of the citation propensity, articles of the sample should be issued within the three-year time-window relating to the paper of interest. Furthermore, when counting the references given by the articles of the sample, one should exclude (1) those references related to publications issued in a different year from that one of the paper of interest and (2) those references to publications not indexed by the database in use. Also, it would be appropriate to exclude papers of different type with respect to the one examined (e.g., when considering a research paper, one should exclude references to reviews, short communications, letters, etc.), due to the different citation propensity.

1. Define a suitable CT_i for estimating the propensity (of the sample's articles) to cite papers issued in the same year of the i-th article of interest. Since citations are counted over a 3-year time-window, three contributions should be taken into account: those related to articles issued respectively 1, 2 and 3 years after the article of interest's issue year. With reference to the sample's articles issued in a specific year, their citation propensity can be estimated by the mean number of references given (of course, excluding some references as explained at point 2). Thus, for an article issued in the year Y, a possible CT_i is given by:

$$CT_i = \bar{r}_{ij(Y+1)} + \bar{r}_{ij(Y+2)} + \bar{r}_{ij(Y+3)} \tag{3}$$

being $\bar{r}_{ij(k)}$ the mean number of references given by the sample's articles issued in the year k.

2. Calculate the *success*-index comparing the citations obtained (c_i) by each (i-th) paper with the respective CT_i value.

Because of the complexity of the calculation of CT_i values, a practical prerequisite of the procedure is to be automated. We built an *ad hoc* software application to query the Scopus database and calculate the indicators of interest automatically. Unfortunately, the formidable amount of database queries for constructing CT_i values makes the calculation relatively time consuming: average time is around 10–15 min for each paper of interest. To overcome this limit—while waiting for database enhancements aimed at reducing response time!—we used a pre-calculated indicator that approaches our *desiderata*, namely the *Database Citation Potential* (*DCP*).

DCP as a practical proxy for CT_i

DCP is used as a normalization term for the *SNIP* indicator, i.e., an annual field-normalized indicator for ranking scientific journals by Moed (2010b). For the purpose of clarity, we briefly mention the main steps for the constructing *DCP*:

- Given a scientific journal J and a reference year Y, we identify a sample of articles issued in that year that cite at least one of the articles published by J, in the 10 years earlier.
- References given by each article of the sample are counted, excluding (1) those references related to sources not indexed by the database in use and (2) those related to articles published outside a 3-year period before Y.
- Finally, *DCP* is given by the average number of references given by the sample's articles.

As a matter of fact, *DCP* estimates the propensity of articles (from the whole literature) issued in the year Y to cite articles homologous to those published by J, issued in the 3 years preceding Y

In our specific analysis, the CT_i of each i-th paper of interest will be given by the *DCP* of the corresponding journal, considering the second year after the issue year as a reference year. This is to make sure that the paper of interest's issue year is in between the three-year

Table 2 Similarities and differences between the estimation of CT_i by (1) the "wished procedure" and (2) the use of DCP_i

Operation	"Wished procedure"	Procedure based on DCP_i
Defining the time-window for counting the citations obtained by the paper of interest	Three consecutive years starting from 1 year after the issue year of the paper of interest	Three consecutive years starting from 1 year after the issue year of the paper of interest
Sample selection		
1) Identification of the papers of the reference sample	Papers citing other papers published by the same journal of the paper of interest (in the 10-year period from the issue year backwards)	Papers citing other papers published by the same journal of the paper of interest (in the 10 years preceding the reference year)
2) Time-window for selecting the articles of the reference sample	Yes	No
3) Distinction between different types of paper (e.g., research paper, review, letter, etc.)	(Citing) papers published during the three consecutive years starting from 1 year after the issue year of the paper of interest	(Citing) papers published during the reference year
4) Time-window for counting the references given by each article of the sample	Issue year of the paper of interest	Three consecutive years preceding the reference year
Definition of a distribution		
1) Distribution used for estimating the citation propensity	References given by the articles of the sample	References given by the articles of the sample
2) Exclusion of references to articles non indexed by the database in use (i.e., Scopus)	Yes	Yes
3) Indicator for estimating CT_i	Sum of the mean number of references to articles issued respectively 1, 2 and 3 years after the issue year of the paper of interest	Mean number of references given by the sample's articles
Computational workload	Very high (more than 15 min for each paper of interest)	Tolerable ($\approx$1/3 min for each paper of interest)

Regarding the procedure based on DCP_i, the "reference year" corresponds to the second year after the issue year of the paper of interest

time-window relating to the *DCP* calculation. Conventionally, the *DCP* related to the *i*-th paper will be referred as DCP_i. Table 2 summarizes the similarities and differences between the construction of CT_i by (1) the "wished procedure" and (2) the DCP_i.

As seen, an important difference between the "wished procedure" and the use of *DCP* is given by the time-window for counting the citations obtained by the sample's papers. In particular, DCP_i is based on the citations given, during the reference year, to papers published one to 3 years before the reference year (Moed 2010a). On the other hand, the "wished procedure" is based on the citations given, during the 3-years after the year of interest, to papers published in that precise year.

Since DCP_i was conceived for normalizing citations accumulated during a single year, while in our specific case a three-year time-window is used, we will conventionally define $CT_i = 2 \cdot DCP_i$.

DCP values related to a particular journal and reference year can be obtained by querying an online application (based on the Scopus database) freely available at (http://www.journalindicators.com, last accessed on July 2012). It is worth remarking that *DCP* is not available for all the journals indexed by Scopus; articles from journals without *DCP* were excluded. This operation should not distort the analysis, since the portion of articles excluded is quite stable—usually around 7–8 %—for the different groups of papers analysed.

The query of Scopus and the free database with *DCP* values was automated by an *ad hoc* application software. The average data-collection time is roughly 20 s for each paper of interest.

We emphasize that the use of *DCP* can greatly simplify and speed up our analysis, without corrupting it: although the *DCP* was conceived by Moed (2010b) to evaluate journals, its use as CT_i at the level of individual publications is not hasty, since *DCP* is an indicator of average citation propensity.

Results

The following subsections describe in detail the results of the three types of analysis.

Comparison among groups of scientists within the same SSD

Table 3 reports some bibliometric indicators concerning the scientific output of the four groups of scientists investigated.

There is some agreement between the *h*-index and the *success*-index values, regardless of the fact that the second indicator implements a field-normalization while the former does not. The reason is that all the scientists belong to the same *SSD*, so it is reasonable to expect that their scientific publications are similar in terms of citation propensity. Of course, large groups are favoured, since the *h*-index and the *success*-index are both size-dependent. For a fair comparison, we introduce *success*/*N* (see the ninth column of Table 3).

We observe the absence of a "critical mass" effect, according to which large groups are likely to be more efficient than small ones, even in terms of size-independent indicators (Merton 1968). The reason may be that scientists of the groups examined are generally split into micro-groups, which tend to work individually.

For all groups, we also note a slight tendency to increase the scientific output over the years, both in terms of relative productivity and relative impact.

Besides, the *success*-indices relating to the overall period (2004–2008) can be determined by simply adding up the relevant annual indices. We reassert that the *h*-index—because of its particular scale of measurement—does not allow this simple composition (Franceschini et al. 2012b).

In conclusion, the four last columns of Table 3 report some statistics concerning the distributions of the DCP_i values related to the papers of the four groups examined. As expected, there are not large differences (see the median values in Table 3). However, we underline that the DCP_i values of group G3 tend to be higher than those of the other groups. This is explained by the fact that this group includes a significant portion of

Table 3 Results of the analysis concerning the comparison between groups of researchers from four Italian universities, within the SSD of *Production Technologies and Manufacturing Systems* (MIUR 2012)

Group	*N*	Period	*P*	*C*	*CPP*	*h*	*Success*	*Success/N*	DCP_i statistics			
									Mean	Median	Sigma	IQR
G1—*Politecnico di Milano*	15	2004	4	24	6.0	4	4	0.27	3.7	3.45	0.76	1.10
		2005	5	18	3.6	3	3	0.20	3.7	3.4	0.63	1.20
		2006	8	27	3.4	3	3	0.20	4.6	4.25	1.50	1.45
		2007	14	50	3.6	4	2	0.13	6.0	5.9	2.02	2.70
		2008	14	56	4.0	5	6	0.40	4.7	4.35	2.12	0.80
		'04–'08	45	175	3.9	7	18	1.20	4.9	4.4	1.92	1.55
G2—*Politecnico di Torino*	20	2004	3	9	3.0	2	2	0.10	5.0	3.4	3.61	5.03
		2005	2	7	3.5	2	1	0.05	4.3	4.25	0.35	0.50
		2006	10	22	2.2	3	2	0.10	4.2	3.75	1.20	1.50
		2007	9	32	3.6	3	2	0.10	4.0	4	0.67	0.48
		2008	8	48	6.0	4	4	0.20	5.4	5.35	1.25	2.15
		'04–'08	32	118	3.7	7	11	0.55	4.5	4.1	1.44	1.35
G3—*Università di Napoli Federico II*	12	2004	11	49	4.5	5	1	0.08	8.5	8.5	3.97	7.18
		2005	17	66	3.9	5	2	0.17	7.5	7.5	2.42	3.30
		2006	13	87	6.7	5	4	0.33	6.5	4.8	4.24	3.73
		2007	20	81	4.1	5	3	0.25	7.2	7.35	3.00	2.40
		2008	14	79	5.6	5	3	0.25	8.4	8.45	3.06	4.50
		'04–'08	75	362	4.8	9	13	1.08	7.6	7.3	3.27	4.28
G4—*Università di Palermo*	13	2004	6	48	8.0	3	2	0.15	7.5	6.6	4.24	5.90
		2005	7	88	12.6	4	3	0.23	7.3	3.7	6.52	7.43
		2006	18	110	6.1	7	13	1.00	3.9	3.8	0.76	0.40
		2007	14	112	8.0	6	10	0.77	4.1	4.2	0.41	0.60
		2008	13	114	8.8	5	6	0.46	6.2	4.6	2.34	4.80
		'04–'08	58	472	8.1	11	34	2.62	5.3	4.2	3.10	0.90

For each group, the following indicators are reported: staff number (*N*), total number of papers (*P*), total number of citations (*C*), average citations per paper (*CPP*), *h*-index of the group (*h*), *success*-index and *success*-index *per capita* (*success/N*). The last four columns report some statistics relating to the distribution of DCP_i values. Indicators are constructed (1) at annual level and (2) considering the whole period of interest (2004–2008). Citations of each paper are counted using a three-year time-window starting from 1 year after the issue year

scientists involved in neighboring disciplines—i.e., Materials Science, Metallurgy and Chemistry—with a significantly larger propensity to the practice of citation. This case is symptomatic of the importance of normalization at the level of sub-discipline (Glänzel et al. 2011; Franceschini and Maisano 2012).

Comparison between university departments

Table 4 reports the results of the comparison between two departments of *Politecnico di Torino*. We observe that, in terms of non-field-normalized indicators (i.e., *C*, *P*, *CPP*, *h*), there is a significant predominance of the second department (D2). The same ranking is confirmed, after introducing the field-normalization through the *success*-index.

However, we note that D2 is favoured by the relatively large *N*. The supremacy of D2 is confirmed even after the size-normalization by *success/N*.

Table 4 Results of the analysis concerning the comparison between groups of researchers from two departments (D1 e D2) of the same university (i.e., *Politecnico di Torino*)

Group	*N*	Period	*P*	*C*	*CPP*	*h*	*Success*	*Success/N*	DCP_i statistics			
									Mean	Median	Sigma	IQR
D1—Dept. of Mathematics	60	2004	26	121	4.7	6	5	0.08	5.9	4.1	4.68	4.00
		2005	37	157	4.2	6	6	0.10	4.2	3.3	2.57	1.43
		2006	36	190	5.3	7	8	0.13	5.4	4.1	3.11	3.30
		2007	38	318	8.4	9	14	0.23	4.6	4.25	1.62	2.20
		2008	50	334	6.7	9	15	0.25	4.9	4.35	1.81	2.10
		‘04–‘08	187	1,120	6.0	17	48	0.80	4.9	4.1	2.78	2.40
D2—Dept. of Chemistry	88	2004	59	307	5.2	10	12	0.14	8.6	8.3	2.72	2.55
		2005	65	548	8.4	12	19	0.22	8.9	8.7	3.08	3.15
		2006	93	705	7.6	13	30	0.34	8.1	7.8	2.34	2.90
		2007	99	868	8.8	16	32	0.36	9.6	9.4	2.61	3.40
		2008	131	1,174	9.0	17	56	0.64	9.1	9.1	2.64	3.58
		‘04–‘08	447	3,602	8.1	23	149	1.69	8.9	8.7	2.69	3.10

For each group, the following indicators are reported: staff number (*N*), total number of papers (*P*), total number of citations (*C*), average citations per paper (*CPP*), *h*-index of the group (*h*), *success*-index and *success*-index *per capita* (*success/N*). The last 4 columns report some statistics relating to the distribution of DCP_i values. Indicators are constructed (1) at annual level and (2) considering the whole period of interest (2004–2008). Citations of each paper are counted using a 3-year time-window starting from 1 year after the issue year

It is also interesting to observe how DCP_i values are distributed; Fig. 2 and the last four columns of Table 4 illustrate the corresponding distributions and statistics. Not surprisingly, the citation propensity of the D2′s papers tend to be much higher than that of D1 (e.g., the median is almost double).

The relative ranking of the two departments looks quite stable over time.

Comparison among whole universities

Table 5 shows the bibliometric indicators relating to the total scientific production of three Italian technical universities, in five consecutive years. The total production of U2 is significantly higher than the others’, due to the larger *N*.

There is a general agreement between the *success*-index and other non-field-normalized indicators because of the fact that the universities examined are quite similar in terms of scientific disciplines covered. In other words, although each university has many scientists specialized in several (sub-)disciplines, the relative diversification of the staff is not so different from university to university.

As regards *success/N*, U3 slightly prevails to U2 and U1. The gap between U1 and U2 is not tremendous, although the staff number of the two universities is much different. The result of this very limited study is somehow consistent with what observed by Abramo et al. (2011), i.e., as regards the level of scientific excellence of the academic staff of Italian universities, there are not strongly “dominant” universities, with an impressive concentration of “star scientists” (like in some important American universities). Instead, “star scientists” are quite uniformly distributed among universities.

Finally, it is interesting to examine the distribution of the DCP_i values concerning the publication output of the three universities (see Fig. 3). As expected, the three distributions

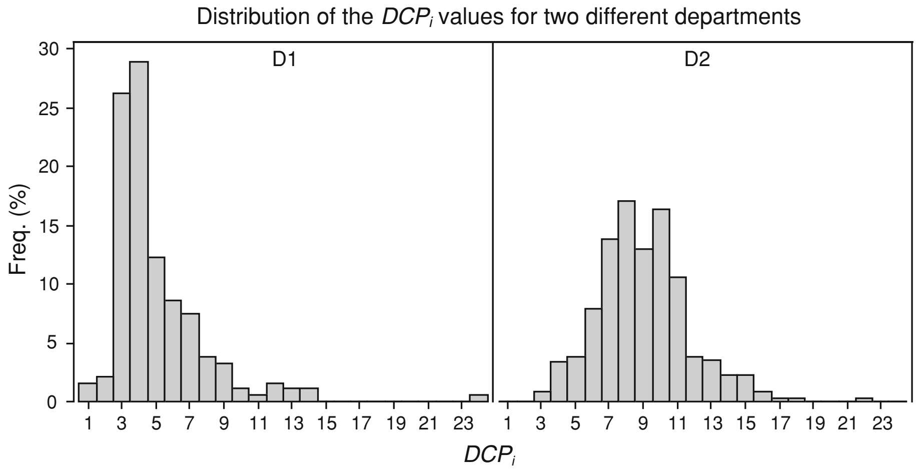

Fig. 2 Distribution of the DCP_i values relating to the publications of the two departments of interest, in the whole period from 2004 to 2008

are quite overlapped. This is quite reasonable since—as seen before—the disciplinary diversification of the staff is relatively similar.

We remark that the use of the *success*-index allows to manage the considerable heterogeneity among scientists, in terms of (sub-)discipline of interest, e.g., in this case, chemistry, physics, mathematics, economics, electronics, etc. Field-normalization avoids the risk of distorting the comparison, that is to say giving more weight to scientific articles from some disciplines while underestimating articles from other disciplines.

Since the normalization implemented by *success*-index concerns the citation impact, we are aware that this indicator may be biased by the different publication rates of (sub-)disciplines. This problem is common to most of the field-normalized indicators and could be solved by introducing a further normalization, at the cost of complicating the indicator considerably.

Sensitivity analysis

Regardless of the choice of how to calculate the CT_i, the *success*-index is as robust as h: in fact, it is not influenced by (1) "big hits", i.e., those papers with an extraordinarily high number of citations, or (2) poorly cited (or even non-cited) articles. We are aware that the first of the two previous aspects can be also seen as a limitation. Specifically, the fact that it does not take into account the so-called "excess citations"—that is to say, the portion of citations above the CT_i threshold—leads to equate papers of medium impact (i.e., with a number of citations just above CT_i) with overcited articles. Similar considerations apply to the h-index.

A partial solution to this problem is given by sensitivity analysis. We clarify this by an example. Suppose to compare two portfolios of scientific articles (such as those representing the output of two scientists or two research institutions) by means of the relevant *success*-indices. CT_i values related to the papers of interest can be calculated according to the procedure based on DCP.

Table 5 Results of the analysis concerning the comparison between three Italian technical universities

Group	*N*	Period	*P*	*C*	*CPP*	*h*	*Success*	*Success/N*	DCP_i statistics			
									Mean	Median	Sigma	IQR
U1—Politecnico di Bari	316	2004	112	525	4.7	10	34	0.11	5.8	4.7	3.47	4.00
		2005	121	772	6.4	14	43	0.14	6.1	5.2	2.98	4.10
		2006	173	1,183	6.8	16	60	0.19	6.1	5.1	3.09	3.83
		2007	149	1,371	9.2	19	63	0.20	6.1	5.1	3.14	3.33
		2008	169	1,174	6.9	16	72	0.23	6.7	5.5	3.66	3.25
		‘04–‘08	724	5,025	6.9	28	272	0.86	6.2	5.3	3.29	3.60
U2—Politecnico di Milano	1,388	2004	483	3,292	6.8	25	166	0.12	7.1	6.6	3.93	4.90
		2005	634	4,141	6.5	24	232	0.17	7.3	6.8	3.91	4.90
		2006	765	5,408	7.1	25	287	0.21	6.9	5.9	3.74	4.50
		2007	843	5,963	7.1	29	276	0.20	7.1	6.5	3.67	4.60
		2008	915	7,607	8.3	33	356	0.26	7.3	6.4	4.73	4.20
		‘04–‘08	3,640	26,411	7.3	45	1,317	0.95	7.1	6.4	4.05	4.60
U3—Politecnico di Torino	837	2004	414	2,608	6.3	23	142	0.17	6.5	5.9	3.24	4.90
		2005	415	3,078	7.4	25	138	0.16	6.9	6.4	6.45	4.88
		2006	548	3,271	6.0	22	185	0.22	6.5	6.2	3.02	4.40
		2007	564	4,471	7.9	25	220	0.26	6.7	5.9	3.23	4.65
		2008	699	5,028	7.2	26	267	0.32	6.9	6.3	4.20	4.20
		‘04–‘08	2,640	18,456	7.0	39	952	1.14	6.7	6.1	4.12	4.50

For each group of researchers the following indicators are reported: staff number (*N*), total number of papers (*P*), total number of citations (*C*), average citations per paper (*CPP*) *h*-index of the group (*h*), *success*-index and *success*-index *per capita* (*success/N*). The last 4 columns report some statistics relating to the distribution of DCP_i values. Indicators are constructed (1) at annual level and (2) considering the whole period of interest (2004–2008). Citations of each paper are counted using a three-year time-window starting from 1 year after the issue year

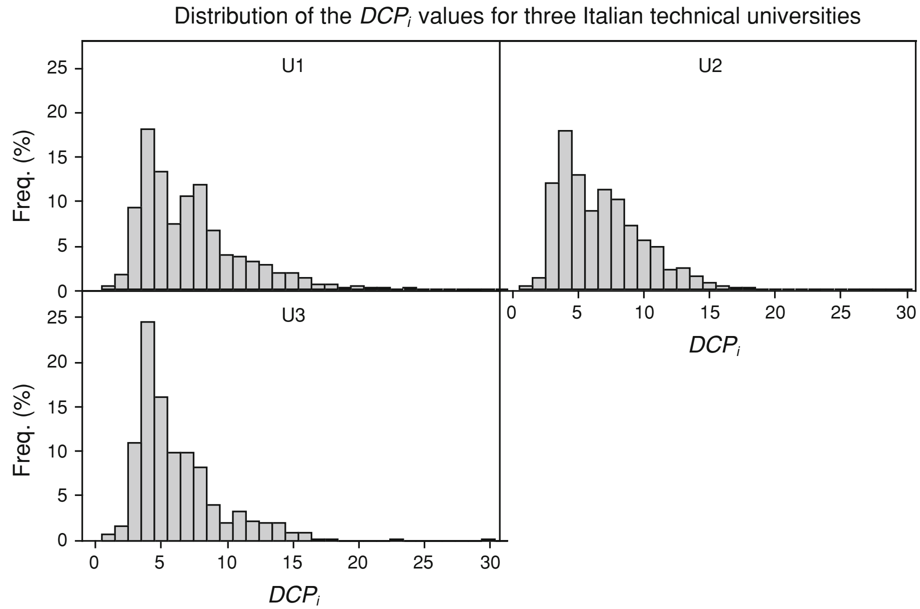

Fig. 3 Distribution of the DCP_i values relating to the publications of the three universities of interest, in the period from 2004 to 2008

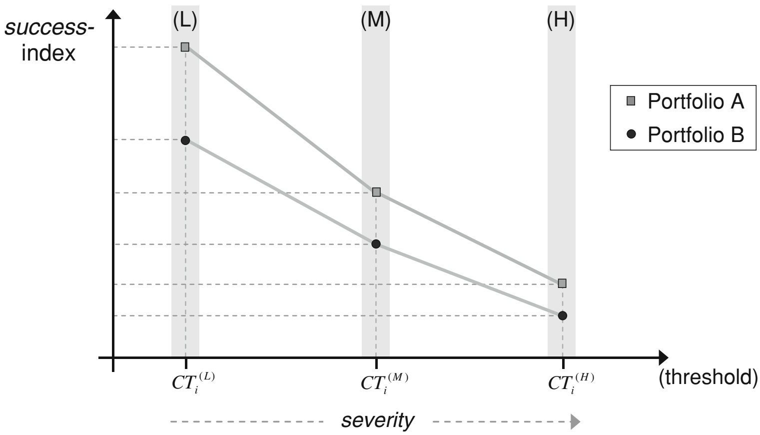

Fig. 4 Profiles resulting from the sensitivity analysis concerning the *success*-indices of two fictitious portfolios of papers (A and B). For each paper, the *success*-index is calculated by varying CT_i on three levels: (i) "permissive" with $CT_i^{(L)}$, (ii) "intermediate" with $CT_i^{(M)}$, and (iii) "severe" with $CT_i^{(H)}$

The procedure can be repeated defining (1) a more "permissive" CT_i threshold—namely $CT_i^{(L)}$—and (2) a more "severe" one—namely $CT_i^{(H)}$. Superscripts "L" and "H" stand for "Low" and "High" respectively. The calculation of the *success*-index can be then repeated using the two new thresholds, producing two new *success*-indices. The

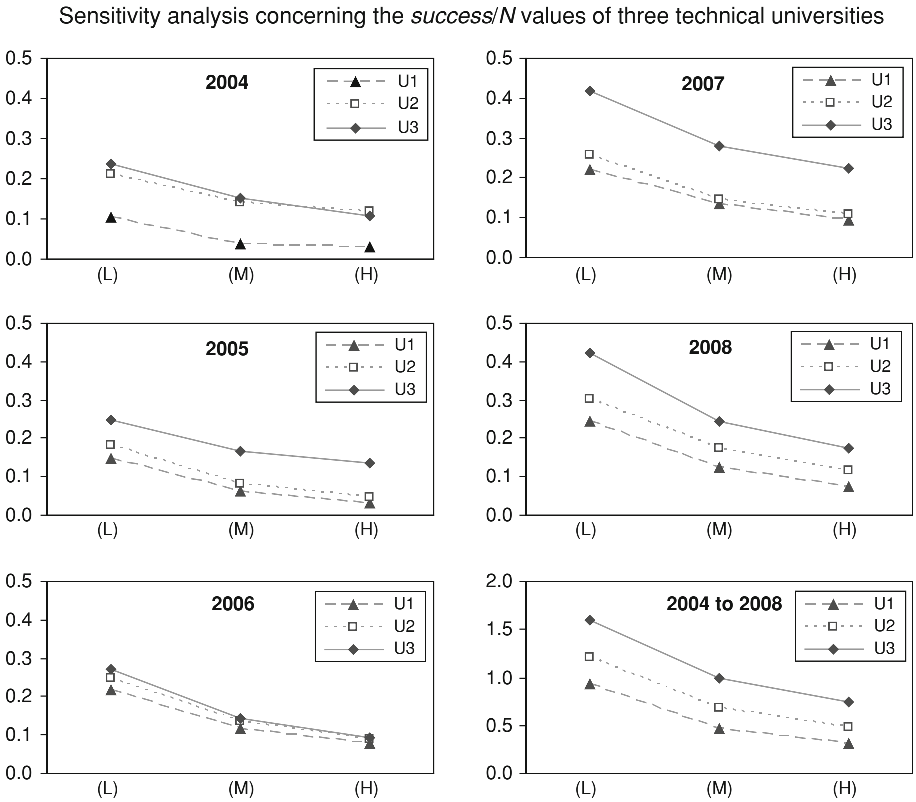

Fig. 5 Profiles resulting from the sensitivity analysis concerning the comparison among the three main Italian technical universities, in terms of *success/N*. Three different CT_i thresholds are used for calculating the *success*-index values (respectively $CT_i^{(L)} = DCP_i$, 464−2621 and $CT_i^{(H)} = 3 \cdot DCP_i$)

overall results of this sensitivity analysis can be represented by means of profiles, such as those represented in Fig. 4.

For the purpose of example, a sensitivity analysis was performed in the case of the comparison among the three technical universities (in "Comparison among whole universities"). The conventional thresholds $CT_i^{(L)} = DCP_i$, $CT_i^{(M)} = 2 \cdot DCP_i$ and $CT_i^{(H)} = 3 \cdot DCP_i$ were used.

Figure 5 reports the resulting profiles, which confirm a certain robustness of the *success*-index: U3 generally prevails over U2, which prevails over U1 and only in one case the relative rankings of U2 and U3 are transposed (precisely, in the year 2004).

Similar considerations apply to the comparisons among groups of researchers and departments.

Concluding remarks

The paper showed—through several practical applications—the relative simplicity and flexibility of the *success*-index when comparing groups of researchers from different disciplines. In particular, it is shown that, apart from a field-normalization at the level of the individual paper, this indicator makes it possible to achieve: (1) normalizations on the basis of the staff number or other "resources" considered, (2) compositions of the results related to multiple years, by a simple sum of the annual contributions.

The robustness of the indicator was tested by a sensitivity analysis. In this sense, it was found that the *success*-index is quite robust. Sensitivity analysis can be also used for investigating whether the major contributions of a group of articles come from the papers of medium, medium–high or high impact.

In the future, we will extend the application of the *success*-index to a large-scale analysis of all the Italian universities (public and private).

References

Abramo, G., D'Angelo, C. A., & Di Costa, F. (2011). National research assessment exercises: a comparison of peer review and bibliometrics rankings. *Scientometrics, 89*(3), 929–941.

Albarrán, P., & Ruiz-Castillo, J. (2011). References made and citations received by scientific articles. *Journal of the American Society for Information Science and Technology, 62*(1), 40–49.

Burrell, Q. L. (2007). On the h-index, the size of the Hirsch core and Jin's A-index. *Journal of Informetrics, 1*(2), 170–177.

Courtault, J. M., & Hayek, N. (2008). On the Robustness of the h-index: a mathematical approach. *Economics Bulletin, 3*(78), 1–9.

Franceschini, F., Galetto, M., & Maisano, D. (2007). *Management by measurement: designing key indicators and performance measurements*. Berlin: Springer.

Franceschini, F., Galetto, M., Maisano, D., & Mastrogiacomo, L. (2012a). The success-index: an alternative approach to the h-index for evaluating an individual's research output. *Scientometrics, 92*(3), 621–641.

Franceschini, F., Galetto, M., Maisano, D., & Mastrogiacomo, L. (2012b). Further clarifications about the *success*-index. *Journal of Informetrics, 6*(4), 669–673.

Franceschini, F., Galetto, M., Maisano, D., & Mastrogiacomo, L. (2012c). An informetric model for the *success*-index. *Journal of Informetrics*.

Franceschini, F., & Maisano, D. (2010). Analysis of the Hirsch index's operational properties. *European Journal of Operational Research, 203*(2), 494–504.

Franceschini, F., & Maisano, D. (2011). Structured evaluation of the scientific output of academic research groups by recent h-based indicators. *Journal of Informetrics, 5*(1), 64–74.

Franceschini, F., & Maisano, D. (2012). Sub-field normalization within the IEEE publication galaxy: a novel approach based on the link between journals and Technical Societies. Proceeding of the 17th international conference on science and technology indicators (STI 2012), 6–8 September 2012, Montréal, Canada.

Garfield, E. (1979). *Citation indexing*. Its theory and application in science, technology and humanities. New York: Wiley.

Glänzel, W. (2011). The application of characteristic scores and scales to the evaluation and ranking of scientific journals. *Journal of Information Science, 37*(1), 40–48.

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.

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.

ISI Web of Knowledge (2012). Essential science indicators. http://thomsonreuters.com (11 July 2012).

Leydesdorff, L., & Opthof, T. (2010). Normalization at the field level: fractional counting of citations. *Journal of Informetrics, 4*(4), 644–646.

Merton, R. K. (1968). The Matthew effect in science. *Science, 159*(3810), 56–63.

MIUR—Ministero dell'Istruzione dell'Università e della Ricerca (2012) http://cercauniversita.cineca.it/php5/settori/index.php.

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.

Radicchi, F., & Castellano, C. (2012). Testing the fairness of citation indicators for comparison across scientific domains: the case of fractional citation counts. *Journal of Informetrics, 6*(1), 121–130.

Scopus—Elsevier (2012) http://www.info.scopus.com (20 Feb 2012).

Vinkler, P. (2004). Characterization of the impact of sets of scientific papers: the Garfield (impact) factor. *Journal of the American Society for Information Science and Technology, 55*(5), 431–435.

Vinkler, P. (2011). Application of the distribution of citations among publications in scientometric evaluations. *Journal of the American Society for Information Science and Technology, 62*(10), 1963–1978.

Waltman, L., Yan, E., & van Eck, N. J. (2011). A recursive field-normalized bibliometric performance indicator: an application to the field of library and information science. *Scientometrics, 89*(1), 301–314.

Wang, X., Liu, D., Ding, K., & Wang, X. (2012). Science funding and research output: a study on 10 countries. *Scientometrics, 91*(2), 591–599.

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.