The MIT Press

RESEARCH ARTICLE

Informed peer review for publication assessments: Are improved impact measures worth the hassle?

Giovanni Abramo[1] iD **, Ciriaco Andrea D'Angelo[2]** iD **, and Giovanni Felici[3]** iD

[1]Laboratory for Studies in Research Evaluation, Institute for System Analysis and Computer Science (IASI-CNR), National Research Council, Rome, Italy
[2]University of Rome "Tor Vergata," Dept of Engineering and Management, Rome, Italy
[3]Institute for System Analysis and Computer Science (IASI-CNR), National Research Council, Rome, Italy

ABSTRACT

In this work we ask whether and to what extent applying a predictor of a publication's impact that is better than early citations has an effect on the assessment of the research performance of individual scientists. Specifically, we measure the total impact of Italian professors in the sciences and economics over time, valuing their publications first by early citations and then by a weighted combination of early citations and the impact factor of the hosting journal. As expected, the scores and ranks of the two indicators show a very strong correlation, but significant shifts occur in many fields, mainly in economics and statistics, and mathematics and computer science. The higher the share of uncited professors in a field and the shorter the citation time window, the more recommendable is recourse to the above combination.

1. INTRODUCTION

Evaluative scientometrics is mainly aimed at measuring and comparing the research performance of entities. In general, a research entity is said to perform better than another if, all production factors being equal, its total output has higher impact. The question then is how to measure the impact of output. Citation-based indicators are more apt to assess scholarly impact than social impact, although it is reasonable to expect that a certain correlation between scholarly and social impact exists (Abramo, 2018).

As far as scholarly impact is concerned, three approaches are available to assess the impact of publications: human judgment (peer review); the use of citation-based indicators (bibliometrics); or drawing on both, whereby bibliometrics informs peer review judgment (informed peer review).

The axiom underlying citation-based indicators is that when a publication is cited, it has contributed to (has had an impact on) the new knowledge encoded in the citing publications—normative theory (Bornmann & Daniel, 2008; Kaplan, 1965; Merton, 1973). There are strong distinctions and objections to the above axiom argued by the social constructivism school, holding that that citing to give credit is the exception, while persuasion is the major motivation for citing (Bloor, 1976; Brooks, 1985, 1986; Gilbert, 1977; Latour, 1987; MacRoberts & MacRoberts, 1984, 1987, 1988, 1989a, 1989b, 1996, 2018; Mulkay, 1976; Teplitskiy, Dueder, et al., 2019).

Although scientometricians, as a shorthand, say that they "measure" scholarly impact, what they actually do is "predict" impact. The reason is that to serve its purpose, any research assessment aimed at informing policy and management decisions cannot wait for the publications life

cycle to be completed (i.e., the publications stop being cited), which may take decades (Song, Situ, et al., 2018; Teixeira, Vieira, & Abreu, 2017; van Raan, 2004).

As a consequence, scientometricians count early citations, not overall citations. The question then is how long should the citation time window be in order for the early citations to be considered an accurate and robust proxy of overall scholarly impact. The longer the citation time window, the more accurate the prediction. In the end, the answer is subjective, because of the embedded trade-off: The appropriate choice of citation time window is a compromise between the two objectives of accuracy and timeliness in measurement, and the relative solutions differ from one discipline to another. The topic has been extensively examined in the literature (Abramo, Cicero, & D'Angelo, 2011; Adams, 2005; Glänzel, Schlemmer, & Thijs, 2003; Nederhof, Van Leeuwen, & Clancy, 2012; Onodera, 2016; Rousseau, 1988; Stringer, Sales-Pardo, & Amaral, 2008; Wang, 2013).

Most studies in evaluative scientometrics focus on providing new creative solutions to the problem of how to best support the measurement of research performance. An extraordinary number of performance indicators continue to be proposed. It suffices to say that at the recent 17th International Society of Scientometrics and Informetrics Conference (ISSI 2019), a special plenary session and five parallel sessions, including 25 contributions altogether (leaving aside poster presentations), were devoted to "novel bibliometric indicators."

Far fewer studies have tackled the problem of how to improve the impact prediction power of early citations, given the inevitable citation short time windows. A number of scholars have proposed combining citation counts with other independent variables related to the publication. Whatever the combination, there is a common awareness that it cannot be the same across disciplines, because the citation accumulation speed and distribution curves vary across disciplines (Baumgartner & Leydesdorff, 2014; Garfield, 1972; Mingers, 2008; Wang, 2013).

It has been shown that in mathematics (and with weaker evidence in biology and earth sciences), for citation windows of 2 years or less the journal's 2-year impact factor (*IF*) is a better predictor of long-term impact than early citations are (Abramo, D'Angelo, & Di Costa, 2010). In every science discipline apart from mathematics, for citation windows of 0 or 1 year only a combination of *IF* and citations is recommended (Bornmann, Leydesdorff, & Wang 2014; Levitt & Thelwall, 2011). The same seems to be valid in the social sciences as well (Stern, 2014). A model based on *IF* and citations to predict long-term citations was proposed by Stegehuis, Litvak, and Waltman (2015). The weighted combination of citations and journal metric percentiles adopted in the Italian research assessment exercise, VQR 2011–2014 (Anfossi, Ciolfi, et al., 2016), proved to be a worse predictor of future impact than citations only (Abramo & D'Angelo, 2016).

To provide practitioners and decision makers with a better predictor of overall impact, and awareness of how the predicting power varies with the citation time window, Abramo, D'Angelo, and Felici (2019) made available, in each of the 170 subject categories (SCs) in the sciences and economics with more than 100 Italian 2004–2006 publications: (a) the weighted combinations of 2-year *IF* and citations, as a function of the citation time window, which best predict overall impact; and (b) the predictive power of each combination.

It emerged that the *IF* has a nonnegligible role only with very short citation time windows (0–2 years); for longer ones, the weight of early citations dominates and the *IF* is not informative in explaining the difference between long-term and short-term citations.

The calibration of the weights by citation time window and SC, and the measurement of the impact indicator, are not as straightforward as the simple measurement of normalized citations.

In this study, we want to find out whether the extra work involved in improving the predicting power of early citations is worthwhile. We ask whether and to what extent applying a predictor of overall impact that is more accurate than early citations has an effect on the research performance ranks of individuals. In this specific case, as a performance indicator we refer to the total impact of individuals. This indicator is particularly appropriate if one needs to identify the top experts in a particular field, for consultancy work or the like. Using an authors' name disambiguation algorithm for Italian academics, we measure the total impact of Italian professors (assistant, associate, and full) in the sciences and in economics over, valuing their publications first by the early citations and then by the weighted citation-*IF* combination provided by Abramo et al. (2019). At this point, we can analyze the extent of variations in rank of individuals in each discipline and field in which they are classified[1].

The rest of the paper is organized as follows. In Section 2, we present the data and method. In Section 3, we report the comparison of the rankings by the two methods of valuing overall impact at field and discipline level. The discussion of results in Section 4 concludes the work.

2. DATA AND METHODS

For the purpose of this study, we are interested in how a different measure of impact affects the ranking of Italian professors by total impact in 2015–2017.

Data on the faculty at each university were extracted from the database of Italian university personnel maintained by the Ministry of Universities and Research (MUR). For each professor this database provides information on their gender, affiliation, field classification, and academic rank at the end of each year[2]. In the Italian university system all academics are classified in one and only one field, a named scientific disciplinary sector (SDS), of which there are 370. SDSs are grouped into 14 disciplines, named *university disciplinary areas* (UDAs).

Data on output and relevant citations are extracted from the Italian Observatory of Public Research, a database developed and maintained by Abramo and D'Angelo, and derived under license from the Clarivate Analytics Web of Science (WoS) Core Collection. Beginning with the raw data of the WoS, and applying a complex algorithm to reconcile the authors' affiliations and disambiguation of the true identity of the authors, each publication (article, letter, review, and conference proceeding) is attributed to the university professor who produced it[3]. Thanks to this algorithm, we can produce rankings by total impact at the individual level on a national scale. Based on the value of *total impact* we obtain a ranking list expressed on a percentile scale of 0–100 (worst to best) of all Italian academics of the same academic rank and SDS.

We limit our field of analysis to the sciences and economics, where the WoS coverage is acceptable for bibliometric assessment. The data set thus formed consists of 38,456 professors from 11 UDAs (mathematics and computer sciences, physics, chemistry, earth sciences, biology, medicine, agricultural and veterinary sciences, civil engineering, industrial and information engineering, psychology, and economics and statistics) and 218 SDSs, as shown in Table 1. Nine point three percent of professors are unproductive (0 publications), and as a consequence their scores but not necessarily their ranks remain unchanged by the two indicators. In fact, the scores and ranks of uncited productive professors (4.2% in all) will change (because *IF* is always

---

[1] To accomplish the assignment, we first need to integrate the *IF*-citation combinations calculated in the 170 SCs with those in the other SCs where the population under observation publishes.

[2] http://cercauniversita.cineca.it/php5/docenti/cerca.php, accessed on March 31, 2020.

[3] The harmonic average of precision and recall (F-measure) of authorships, as disambiguated by the algorithm, is around 97% (2% margin of error, 98% confidence interval).

Table 1. Data set of the analysis. Italian professors holding formal faculty roles for at least 2 years over the 2015–2017 period, by UDA and academic rank

UDA*	No. of SDSs	Total professors	Unproductive	Uncited productive
1	10	3019	380 (12.6%)	227 (7.5%)
2	8	2146	103 (4.8%)	42 (2.0%)
3	11	2815	59 (2.1%)	23 (0.8%)
4	12	1010	50 (5.0%)	11 (1.1%)
5	19	4630	184 (4.0%)	53 (1.1%)
6	50	9159	748 (8.2%)	231 (2.5%)
7	30	2948	190 (6.4%)	76 (2.6%)
8	9	1500	129 (8.6%)	63 (4.2%)
9	42	5290	246 (4.7%)	169 (3.2%)
10	10	1402	168 (12.0%)	68 (4.9%)
11	17	4537	1312 (28.9%)	635 (14.0%)
Total	218	38456	3569 (9.3%)	1598 (4.2%)

* 1: Mathematics and computer science; 2: Physics; 3: Chemistry; 4: Earth sciences; 5: Biology; 6: Medicine; 7: Agricultural and veterinary sciences; 8: Civil engineering; 9: Industrial and information engineering; 10: Psychology; 11: Economics and statistics.

above 0). Measuring the latter's impact by citations only, their score (0) and rank would be the same as for unproductive professors, But it would not when measured by the weighted combination of normalized citations and *IF*.

We measure impact in two ways: One values publications by early citations only, and the other by the weighted combinations of citations and *IF*[4], as a function of the citation time window and field of research, which best predict future impact (Abramo et al., 2019).

Because citation behavior varies across fields, we standardize the citations for each publication with respect to the average of the distribution of citations for all publications indexed in the same year and the same SC[5]. We apply the same procedure to the *IF*.

Furthermore, research projects frequently involve a team of scientists, which is registered in the coauthorship of publications. In this case, we account for the fractional contributions of scientists to outputs, which is sometimes further signaled by the position of the authors in the list of authors.

The yearly total impact of a professor, termed *TI*, is then defined as

$$TI = \frac{1}{t}\sum_{i=1}^{N} c_i f_i,$$

where t is the number of years on staff of professor during the observation period; N is the number of publications by the professor in the period under observation; c_i is alternatively

---

[4] The journal *IF* refers to the year of publication.

[5] Abramo, Cicero, and D'Angelo (2012) demonstrated that the average of the distribution of citations received for all cited publications of the same year and SC is the best-performing scaling factor.

(a) citations received by publication i normalized to the average of distribution of citations received for all cited publications in the same year and SC of publication i or (b) weighted combination of normalized citations and normalized *IF* of the hosting journal, whereby weights differ across citation time windows and SCs, as in Abramo et al. (2019); and f_i is the fractional contribution of the professor to publication i.

The fractional contribution equals the inverse of the number of authors in those fields where the practice is to place the authors in simple alphabetical order, but assumes different weights in other cases. For the life sciences, the widespread practice in Italy is for the authors to indicate the various contributions to the published research by the order of names in the listing of the authors. For the life sciences, we give different weights to each coauthor according to position in the list of authors and the character of the coauthorship (intramural or extramural).[6]

For reasons of significance, the analysis is limited to those professors who held formal faculty roles for at least 2 years over the 2015–2017 period.

Citations are observed at December 31, 2018, implying citation time windows ranging from 1–4 years.

3. RESULTS

In the following, we present the score and rank of performance by total impact of Italian professors, by SDS and UDA, as measured respectively by

- early citations (TI_C)
- the weighted combination of citations and *IF* of the hosting journal (TI_{WC})

As already noted, no variations will occur for professors with no publications in the period under observation. We expect instead significant variations in score and rank for professors with uncited publications. In fact, while TI_C is zero, TI_{WC} will be above zero.

As an example, Table 2 shows the scores and ranks by TI_C and TI_{WC} for the 26 Italian professors in the SDS Aerospace propulsion. The score variation is zero for the two unproductive professors at the bottom of the list, while it is a maximum for the uncited productive professors (ID 49113 and 2592). Twelve professors experience no shift, among them the top five in ranking. A few pairs swap positions (e.g., ID 78162 and ID 49106). The maximum shift is three positions.

The SDS Industrial chemistry consists of 114 professors, mostly productive and cited. Figure 1 shows the dispersion of their impact. The very strong correlation of scores (Pearson $\rho = 0.999$) and ranks (Spearman $\rho = 0.998$) by TI_C and TI_{WC} are as expected.

Higher dispersion (Figure 2) occurs instead for the 73 professors in the SDS Complementary mathematics, whereby about two-thirds (50) of professors present zero TI_C, and 20% (15), while productive, are uncited (TI_C above 0). As a matter of fact, noticeable shifts in relative scores occur for high performers too (top right of the diagram), notwithstanding a very strong score

---

6   If the first and last authors belong to the same university, 40% of the citation is attributed to each of them; the remaining 20% is divided among all other authors. If the first two and last two authors belong to different universities, 30% of the citation is attributed to the first and last authors, 15% of the citation is attributed to the second and last authors but one, and the remaining 10% is divided among all others. The weightings were assigned following advice from senior Italian professors in the life sciences. The values could be changed to suit different practices in other national contexts.

Table 2. Ranking lists by total impact (TI_C and TI_{WC}) of Italian professors in the SDS Aerospace propulsion

ID	TI_C			TI_{WC}			Δ score	Δ rank	
	Score	Rank	Percentile	Score	Rank	Percentile			
10712	2.703	1	100	3.360	1	100.0	24.3%	0	=
49114	0.824	2	96	1.268	2	96.0	53.9%	0	=
49109	0.773	3	92	0.906	3	92.0	17.2%	0	=
4045	0.666	4	88	0.853	4	88.0	28.2%	0	=
2590	0.633	5	84	0.759	5	84.0	19.8%	0	=
78162	0.548	6	80	0.698	7	76.0	27.5%	1	↓
49106	0.504	7	76	0.731	6	80.0	44.9%	1	↑
4047	0.365	8	72	0.489	8	72.0	34.0%	0	=
37761	0.240	9	68	0.383	10	64.0	59.4%	1	↓
4044	0.224	10	64	0.479	9	68.0	113.7%	1	↑
2597	0.211	11	60	0.340	11	60.0	61.4%	0	=
5463	0.191	12	56	0.287	12	56.0	50.7%	0	=
49118	0.183	13	52	0.268	13	52.0	46.7%	0	=
49115	0.105	14	48	0.132	14	48.0	26.1%	0	=
49117	0.074	15	44	0.085	18	32.0	15.0%	3	↓
78159	0.069	16	40	0.103	15	44.0	48.6%	1	↑
2595	0.059	17	36	0.085	17	36.0	43.3%	0	=
4046	0.059	17	36	0.072	19	28.0	21.3%	2	↓
4048	0.047	19	28	0.099	16	40.0	110.6%	3	↑
49111	0.036	20	24	0.038	21	20.0	5.6%	1	↓
2589	0.024	21	20	0.025	22	16.0	5.6%	1	↓
87212	0.020	22	16	0.040	20	24.0	97.5%	2	↑
49113	0.000	23	0	0.012	23	12.0	∞	0	=
2592	0.000	23	0	0.004	24	8.0	∞	1	↓
2599	0.000	23	0	0.000	25	0.0	n.a.	2	↓
40946	0.000	23	0	0.000	25	0.0	n.a.	2	↑

correlation (Pearson $\rho = 0.988$). The ability of TI_{WC} to discriminate the impact of uncited publications, and therefore the relevant performance of uncited professors, explains the lower rank correlation (Spearman $\rho = 0.915$). Although variations in score are not that noticeable, those in rank are. To better show that, Figure 3 reports the share of professors experiencing a rank shift in both SDSs. In Complementary mathematics, over 60% of professors do not change rank (50%
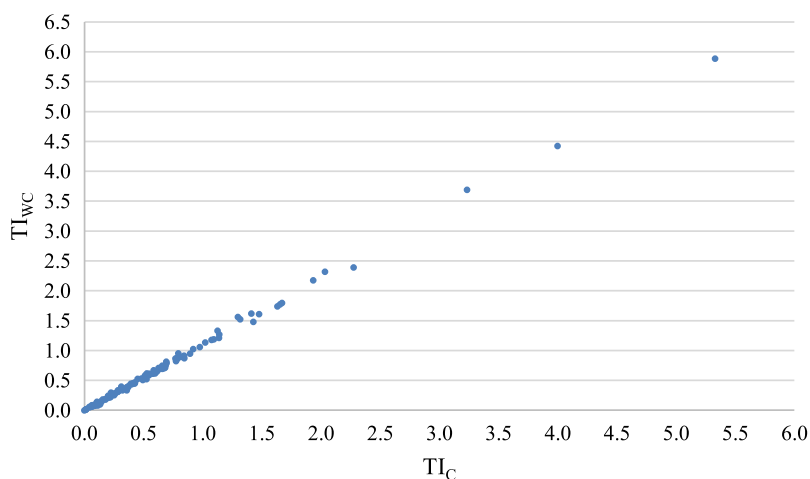
Figure 1. Score dispersion by TI_C and TI_{WC} of the 114 Italian professors in the SDS Industrial chemistry.

could not, as they were unproductive). The remaining 40% present shifts that are in some cases quite noticeable: Five professors improve their rank by no fewer than 10 positions. Rank shifts are less evident in Industrial chemistry: The average shift is 1.47 positions, as compared to 1.89 in Complementary mathematics. In the former SDS, because of the lower number of unproductive professors, shifts affect a higher share of the population, namely 70%.

For a better appreciation of the rank variations in the whole SDS spectrum, Figure 4 shows the box plots of the average percentile shifts in the SDSs of each UDA, while Table 3 presents some relevant descriptive statistics.

Economics and statistics is the UDA with the highest average percentile shift (6.5), the highest dispersion among SDSs (3.6 standard deviations), and the widest range of percentile shift, from 1.4 of SECS-P/13 (Commodity science) to 15.9 of SECS-P/04 (History of economic thought). It is followed by Mathematics and computer science, whose range of variation of the percentile shift is between 2.0 of MAT/08 (Numerical analysis) and 12.6 of MAT/04 (Complementary
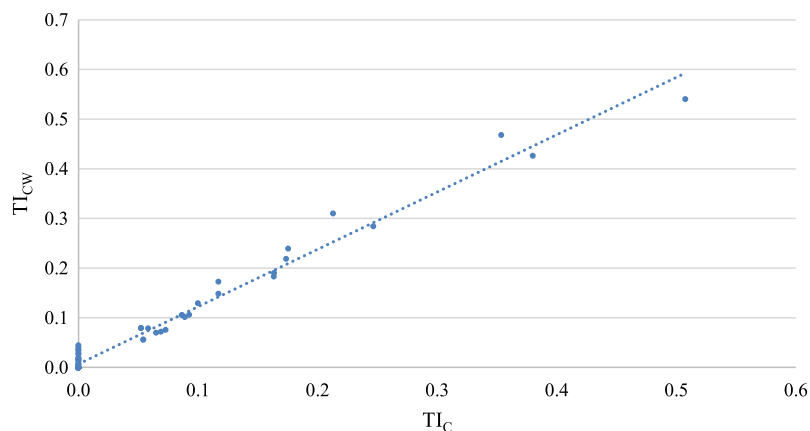

Figure 2. Score dispersion by TI_C and TI_{WC} of the 73 Italian professors in the SDS Complementary mathematics.
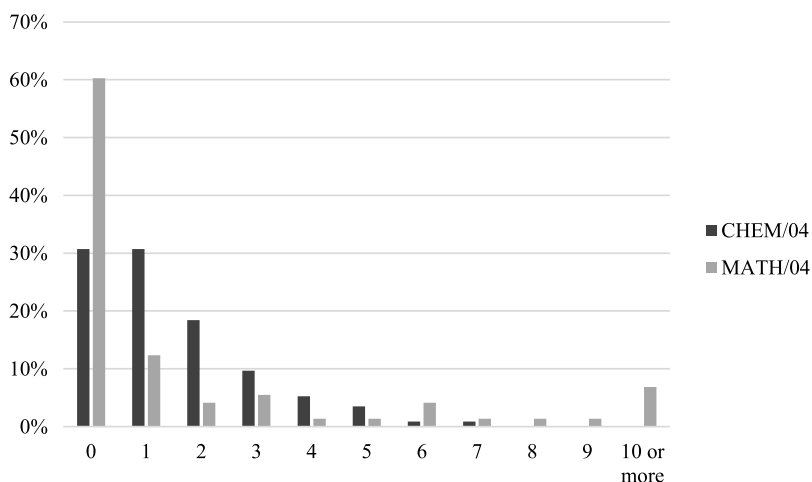
Figure 3. Share of professors experiencing a rank shift in the SDSs Industrial chemistry (CHEM/04) and Complementary mathematics (MATH/04).

mathematics). In contrast, UDAs 4 (Earth sciences) and 5 (Biology) show the lowest dispersion among SDSs (0.3 standard deviation) and quite low average percentile shifts. In UDA Medicine, a peculiar case occurs: In SDS MED/47 (Nursing and midwifery) the two ranking lists are exactly the same. The same occurs also in two SDSs of Industrial and information engineering: ING-IND/29 (Raw materials engineering) and ING-IND/30 (Hydrocarburants and fluids of the subsoil). In general, in 17 out of 218 SDSs, the average percentile shift is never below five percentiles.

In general, the correlation between TI_C and TI_{WC} is very strong. Table 4 presents some descriptive statistics of both Pearson ρ (score) and Spearman ρ (rank) for the SDSs of each UDA.
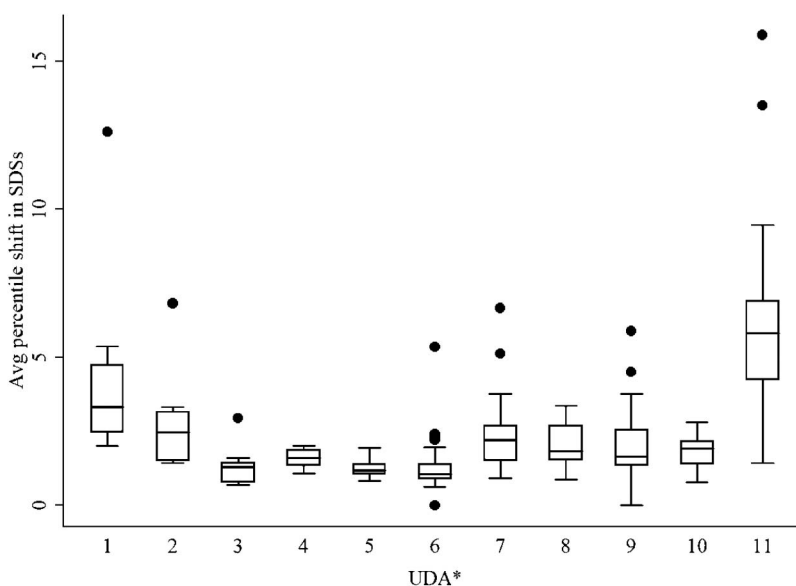

Figure 4. Box plot of average percentile shifts in the SDSs of each UDA. * 1: Mathematics and computer science; 2: Physics; 3: Chemistry; 4: Earth sciences; 5: Biology; 6: Medicine; 7: Agricultural and veterinary sciences; 8: Civil engineering; 9: Industrial and information engineering; 10: Psychology; 11: Economics and statistics.

Table 3. Descriptive statistics of percentile shifts in the SDSs of each UDA

UDA	Min	Max	Avg	St. dev.
1	2.0 (MAT/08)	12.6 (MAT/04)	4.3	3.0
2	1.4 (FIS/03)	6.8 (FIS/08)	2.8	1.7
3	0.7 (CHIM/09)	3.0 (CHIM/12)	1.3	0.6
4	1.1 (GEO/07)	2.0 (GEO/10)	1.6	0.3
5	0.8 (BIO/14)	1.9 (BIO/05)	1.3	0.3
6	0 (MED/47)	5.3 (MED/02)	1.3	0.7
7	0.9 (AGR/16)	6.7 (AGR/06)	2.3	1.2
8	0.9 (ICAR/03)	3.4 (ICAR/06)	2.0	0.7
9	0 (ING-IND/29; ING-IND/30)	5.9 (ING-IND/02)	1.9	1.1
10	0.8 (M-EDF/02)	2.8 (M-PSI/07)	1.8	0.6
11	1.4 (SECS-P/13)	15.9 (SECS-P/04)	6.5	3.6

* AGR/06: Wood technology and forestry operations; AGR/16: Agricultural Microbiology; BIO/05: Zoology; BIO/14: Pharmacology; CHIM/09: Pharmaceutical and technological applications of chemistry; CHIM/12: Chemistry for the environment and for cultural heritage; FIS/03: Physics of matter; FIS/08: Didactics and history of physics; GEO/07: Petrology and petrography; GEO/10: Solid Earth geophysics; ICAR/03: Sanitary and environmental engineering; ICAR/06: Topography and cartography; ING-IND/02: Ship structures and marine engineering; ING-IND/29: Engineering of raw materials; ING-IND/30: Hydrocarbons and underground fluids; MAT/04: Mathematics education and history of mathematics; MAT/08: Numerical analysis; MED/02: Medical history; MED/47: Midwifery; M-EDF/02: Methods and teaching of sports activities; M-PSI/07: Dynamic psychology; SECS-P/04: History of economic thought; SECS-P/13: Commodity science.

Table 4. Descriptive statistics of correlation coefficients for TI_C and TI_{WC} in the SDSs of each UDA

UDA*	Pearson correlation				Spearman correlation			
	Min.	Max.	Avg.	St. dev.	Min.	Max.	Avg.	St. dev.
1	0.986	0.999	0.995	0.004	0.915	0.995	0.981	0.023
2	0.984	0.999	0.994	0.005	0.969	0.997	0.990	0.009
3	0.998	1	0.999	0.001	0.973	0.999	0.996	0.007
4	0.997	1	0.999	0.001	0.994	0.998	0.996	0.001
5	0.998	1	0.999	0.001	0.995	0.999	0.998	0.001
6	0.957	1	0.998	0.006	0.964	1	0.997	0.005
7	0.981	1	0.997	0.004	0.939	0.999	0.992	0.011
8	0.994	0.999	0.998	0.002	0.989	0.999	0.995	0.003
9	0.983	1	0.997	0.003	0.953	1	0.994	0.008
10	0.995	1	0.998	0.002	0.994	0.999	0.996	0.002
11	0.993	0.998	0.996	0.002	0.884	0.997	0.969	0.029

* 1: Mathematics and computer science; 2: Physics; 3: Chemistry; 4: Earth sciences; 5: Biology; 6: Medicine; 7: Agricultural and veterinary sciences; 8: Civil engineering; 9: Industrial and information engineering; 10: Psychology; 11: Economics and statistics.
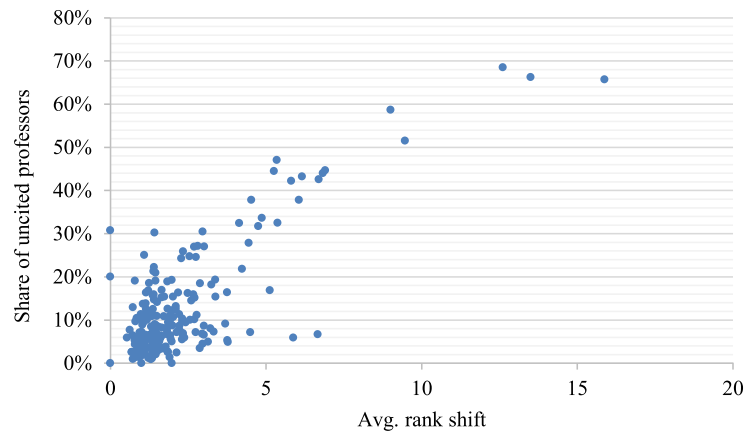
Figure 5. Field dispersion per share of uncited professors and average rank shift by TI_C and TI_{WC}.

As for the scores, the minimum correlation (0.957) occurs in an SDS of Medicine (MED/02 – History of medicine). As for the ranks, the minimum occurs (0.884) in an SDS of Economics and statistics, SECS-P/04 (History of economic thought), which also stands out for the maximum average percentile shift among all SDSs (Table 3). It is a relatively small SDS, 35 professors in all, two-thirds of whom have zero TI_C.

The rank variations in general appear strongly correlated with the share of productive professors with zero TI_C (i.e., with only uncited publications). The correlation between the two variables is shown in Figure 5 (Pearson ρ = 0.791).

A typical way to report performance is by quartile ranking. We then analyze the performance quartile shifts by the two indicators of impact. Table 5 represents the contingency matrix of the performance quartile by TI_C and TI_{WC} of all 38,456 professors of the data set. Because of the strong correlation between the two indicators, we observe an equally strong concentration of frequencies along the main diagonal: In 93.3% of cases (24.52% + 22.90% + 20.96% + 24.94%), the performance quartile remains unchanged. 0.7% of professors in Q1 by TI_C shift to Q2 by TI_{WC} and 0.8% of professors in Q4 by TI_C shift to Q3 by TI_{WC}.

Table 6 shows the shift distributions by UDA. Economics and statistics presents the highest share of professors (17.1%) shifting quartile, followed by Mathematics (9%). In the remaining UDAs shares range between 4% and 7%. It must be noted that 0.4% of professors (154) experience two quartile shifts, and all but two shift from the bottom to above the median. They are mainly in Economics and statistics, and in Mathematics.

Table 5. Professors' performance quartile distribution as measured by TI_C and TI_{WC}

		TI_{WC}			
		I	II	III	IV
TI_C	I	24.52%	0.70%	0.00%	0.00%
	II	0.69%	22.90%	1.03%	0.00%
	III	0.00%	0.96%	20.96%	0.84%
	IV	0.00%	0.40%	2.06%	24.94%

Table 6. Distribution of quartile shifts based on FSS_P as measured by TI_C and TI_{WC}, by UDA (percentage of the total UDA staff in parentheses)

UDA*	Shifting quartiles	Shifting two quartiles
1	271 (9.0%)	15 (0.50%)
2	160 (7.5%)	0 (0%)
3	122 (4.3%)	1 (0.04%)
4	63 (6.2%)	0 (0%)
5	200 (4.3%)	0 (0%)
6	349 (3.8%)	1 (0.01%)
7	197 (6.7%)	1 (0.03%)
8	77 (5.1%)	0 (0%)
9	284 (5.4%)	0 (0%)
10	71 (5.1%)	0 (0%)
11	778 (17.1%)	136 (3.00%)
Total	2572 (6.7%)	154 (0.40%)

* 1: Mathematics and computer science; 2: Physics; 3: Chemistry; 4: Earth sciences; 5: Biology; 6: Medicine; 7: Agricultural and veterinary sciences; 8: Civil engineering; 9: Industrial and information engineering; 10: Psychology; 11: Economics and statistics.

4. CONCLUSIONS

Evaluative scientometrics is mainly aimed at measuring and comparing the research performance of individuals and organizations. A critical issue in the process is the accurate prediction of the scholarly impact of publications when citation short time windows are allotted. This is often the case when the evaluation is geared to informed decision-making.

Better impact prediction accuracy often involves complex, costly, and time consuming measurements. Pragmatism requires an analysis of the effects of improved indicators on the performance ranking of the subjects under evaluation. This study follows up the work by the same authors (Abramo et al., 2019), which demonstrated that especially with very short time windows (0–2 years) the *IF* can be combined with early citations as a powerful covariate for predicting long-term impact.

Using the outcomes of such work (i.e., the weighted combinations of *IF* and citations as a function of the citation time window) that best predict the overall impact of single publications in each SC, we have been able to measure the 2015–2017 total impact of all Italian professors in the sciences and economics, and to analyze the variations in performance ranks when using early citations only.

As expected, scores and ranks by the two indicators show a very strong correlation. Nevertheless, in 7% of SDSs, the average shift never goes below 5 percentiles and is 15.6 and 12.9 on average in the SDSs, respectively, of Economics and statistics and Mathematics and computer science.

In terms of quartile shifts, almost 7% of professors undergo them. In Economics and statistics, 3% of professors shift from Q4 to above the median.

A strong correlation can be seen between the rate of shifts in rank and the share of uncited professors in the SDS. The total impact of uncited professors is in fact nil by TI_C, but above zero by TI_{WC}. In short, TI_{WC} can better discriminate the performance of professors in the left tail of the distribution. The higher the share of uncited professors in an SDS, the more recourse to TI_{WC} is recommended. Furthermore, the shorter the citation time window, the heavier the relative weight of IF in predicting the long-term impact. TI_{WC} is then highly recommended when citation time windows are short and the rate of uncited professors is high.

In the case of national research assessment exercises based on informed peer review or on bibliometrics only, the weighted combination of normalized citations and IF to rank publications might be adopted, as the weights can be made available by the authors for all SCs and citation time windows up to 6 years.

Possible future investigations within this stream of research might concern the effect of the improved indicator of publications' impact on the performance score and rank of research organizations and research units.

AUTHOR CONTRIBUTIONS

Giovanni Abramo: Conceptualization; Formal analysis; Investigation; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing—original draft; Writing—review & editing. Ciriaco Andrea D'Angelo: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Software; Validation; Visualization; Writing—original draft; Writing—review & editing. Giovanni Felici: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Software; Validation; Visualization; Writing—original draft; Writing—review & editing.

COMPETING INTERESTS

FUNDING INFORMATION

DATA AVAILABILITY

The authors use data licensed from Clarivate Analytics. They are not allowed to publish this data.

REFERENCES

Abramo, G. (2018). Revisiting the scientometric conceptualization of impact and its measurement. *Journal of Informetrics*, *12*(3), 590–597.

Abramo, G., Cicero, T., & D'Angelo, C. A. (2011). Assessing the varying level of impact measurement accuracy as a function of the citation window length. *Journal of Informetrics*, *5*(4), 659–667.

Abramo, G., Cicero, T., & D'Angelo, C. A. (2012). Revisiting the scaling of citations for research assessment. *Journal of Informetrics*, *6*(4), 470–479.

Abramo, G., & D'Angelo, C. A. (2016). Refrain from adopting the combination of citation and journal metrics to grade publications, as used in the Italian national research assessment exercise (VQR 2011–2014). *Scientometrics*, *109*(3), 2053–2065.

Abramo, G., D'Angelo, C. A., & Di Costa, F. (2010). Citations versus journal impact factor as proxy of quality: Could the latter ever be preferable? *Scientometrics*, *84*(3), 821–833.

Abramo, G., D'Angelo, C. A., & Felici, G. (2019). Predicting long-term publication impact through a combination of early citations and journal impact factor. *Journal of Informetrics*, *13*(1), 32–49.

Adams, J. (2005). Early citation counts correlate with accumulated impact. *Scientometrics*, *63*(3), 567–581.

Anfossi, A., Ciolfi, A., Costa, F., Parisi, G., & Benedetto, S. (2016). Large-scale assessment of research outputs through a weighted combination of bibliometric indicators. *Scientometrics*, *107*(2), 671–683.

Baumgartner, S., & Leydesdorff, L. (2014). Group-based trajectory modelling (GBTM) of citations in scholarly literature: Dynamic

qualities of "transient" and "sticky" knowledge claims. *Journal of the American Society for Information Science and Technology*, *65*(4), 797–811.

Bloor, D. (1976). *Knowledge and social imagery*. London: Routledge & Kegan Paul.

Bornmann, L., & Daniel, H. D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*(1), 45–80.

Bornmann, L., Leydesdorff, L., & Wang, J. (2014). How to improve the prediction based on citation impact percentiles for years shortly after the publication date? *Journal of Informetrics*, *8*(1), 175–180.

Brooks, T. A. (1985). Private acts and public objects: An investigation of citer motivations. *Journal of the American Society for Information Science*, *36*(4), 223–229.

Brooks, T. A. (1986). Evidence of complex citer motivations. *Journal of the American Society for Information Science*, *37*(3), 34–36.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, *178*, 471–479.

Gilbert, G. N. (1977). Referencing as persuasion. *Social Studies of Science*, *7*(1), 113–122.

Glänzel, W., Schlemmer, B., & Thijs, B. (2003). Better late than never? On the chance to become highly cited only beyond the standard bibliometric time horizon. *Scientometrics*, *58*(3), 571–586.

Kaplan, N. (1965). The norms of citation behavior: Prolegomena to the footnote. *American Documentation*, *16*(3), 179–184.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.

Levitt, J. M., & Thelwall, M. (2011). A combined bibliometric indicator to predict article impact. *Information Processing and Management*, *47*(2), 300–308.

MacRoberts, M. H., & MacRoberts, B. R. (1984). The negational reference: Or the art of dissembling. *Social Studies of Science*, *14*(1), 91–94.

MacRoberts, M. H., & MacRoberts, B. R. (1987). Another test of the normative theory of citing. *Journal of the American Society for Information Science*, *38*(4), 305–306.

MacRoberts, M. H., & MacRoberts, B. R. (1988). Author motivation for not citing influences: A methodological note. *Journal of the American Society for Information Science*, *39*(6), 432–433.

MacRoberts, M. H., & MacRoberts, B. R. (1989a). Citation analysis and the science policy arena. *Trends in Biochemical Science*, *14*(1), 8–10.

MacRoberts, M. H., & MacRoberts, B. R. (1989b). Problems of citation analysis: A critical review. *Journal of the American Society for Information Science*, *40*(5), 342–349.

MacRoberts, M. H., & MacRoberts, B. R. (1996). Problems of citation analysis. *Scientometrics*, *36*(3), 435–444.

MacRoberts, M. H., & MacRoberts, B. R. (2018). The mismeasure of science: Citation analysis. *Journal of the Association for Information Science and Technology*, *69*(3), 474–482.

Merton, R. K. (1973). Priorities in scientific discovery. In R. K. Merton (Ed.), *The sociology of science: Theoretical and empirical investigations* (pp. 286–324). Chicago: University of Chicago Press.

Mingers, J. (2008). Exploring the dynamics of journal citations: Modelling with S-curves. *Journal Operational Research Society*, *59*(8), 1013–1025.

Mulkay, M. (1976). Norms and ideology in science. *Social Science Information*, *15*(4–5), 637–656.

Nederhof, A. J., Van Leeuwen, T. N., & Clancy, P. (2012). Calibration of bibliometric indicators in space exploration research: A comparison of citation impact measurement of the space and ground-based life and physical sciences. *Research Evaluation*, *21*(1), 79–85.

Onodera, N. (2016). Properties of an index of citation durability of an article. *Journal of Informetrics*, *10*(4), 981–1004.

Rousseau, R. (1988). Citation distribution of pure mathematics journals. In L. Egghe & R. Rousseau (Eds.), *Informetrics 87/88* (pp. 249–262). Amsterdam: Elsevier.

Song, Y., Situ, F. L., Zhu, H. J., & Lei, J. Z. (2018). To be the Prince to wake up Sleeping Beauty: The rediscovery of the delayed recognition studies. *Scientometrics*, *117*(1), 9–24.

Stegehuis, C., Litvak, N., & Waltman, L. (2015). Predicting the long-term citation impact of recent publications. *Journal of Informetrics*, *9*(3), 642–657.

Stern, D. I. (2014). High-ranked social science journal articles can be identified from early citation information. *PLOS ONE*, *9*(11), 1–11.

Stringer, M. J., Sales-Pardo, M., & Amaral, L. A. N. (2008). Effectiveness of journal ranking schemes as a tool for locating information. *PLOS ONE*, *3*(2), e1683. https://doi.org/10.1371/journal.pone.0001683

Teixeira, A. A. C., Vieira, P. C., & Abreu, A. P. (2017). Sleeping Beauties and their princes in innovation studies. *Scientometrics*, *110*(2), 541–580.

Teplitskiy, M., Dueder, E., Menietti, M., & Lakhani, K. (2019). Why citations don't mean what we think they mean: Evidence from citers. Proceedings of the 17th International Society of Scientometrics and Informetrics Conference (ISSI 2019). September 2–5, Rome, Italy.

van Raan, A. F. J. (2004). Sleeping beauties in science. *Scientometrics*, *59*(3), 461–466.

Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, *94*(3), 851–872.