Scientometrics
https://doi.org/10.1007/s11192-025-05317-y

Measuring scholarly performance using comprehensive standardized research-teaching (RT) score

Nicola Scafetta[1]

Received: 30 June 2024 / Accepted: 4 April 2025

Abstract
University faculty members and participants in scientific competitions are typically evaluated based on metrics derived from their published works and other relevant academic activities. However, designing a robust mathematical algorithm to process bibliometric information is challenging, and accessible computer codes are often scarce. Consequently, evaluation committees may resort to improvised, mathematically inadequate, poorly standardized, and overly simplistic evaluation methods, which can yield unfair and not-transparent outcomes. This paper introduces a novel algorithm, the "*RT-score*", designed to assess and rank the research and teaching performance of a group of academics. The RT-score builds upon the "*C-score*", which is currently used to generate the "*Stanford/Elsevier World Top 2% Most Influential Scientists*" list. The RT-score incorporates several complementary bibliometric indicators, including productivity (number of publications), impact (citations), and the involvement of the individual researcher in the published works. The RT-score also emphasizes the most recent and impactful publications while incorporating parameters that account for variations in the number of articles and citation density across different scientific disciplines, funding, and other pertinent aspects. Finally, it combines these metrics with a measure of teaching experience and other academic activities. The RT-score aims to address key recommendations from DORA, CoARA, and the Leiden Manifesto concerning research assessment reform. The supplement provides a MATLAB code that implements the proposed algorithm.

Keywords Scholarly performance · Citation metrics · Teaching metrics · Faculty rank

Introduction

The evaluation and ranking of faculty members' research and scholarly achievements in higher education institutions has traditionally relied on metrics derived from their published works and other academic activities. Ranking the scholarly performance of academics is crucial for decisions regarding hiring, promotion, tenure, and rewards such as

✉ Nicola Scafetta
nicola.scafetta@unina.it

[1] Department of Earth Sciences, Environment and Georesources, University of Naples Federico II, Complesso Universitario di Monte S. Angelo, Naples, Italy

funding and recognition (Moher et al., 2018). It is essential to develop robust evaluation algorithms, and universities must ensure transparency in the criteria and methodologies they employ for these evaluations. While it can be argued that scientific assessments should primarily rely on qualitative judgments regarding the impact and scientific significance of research—where peer review plays a pivotal role (CoARA, 2022; Hicks et al., 2015)—these evaluations should also be complemented by the responsible application of various quantitative indicators.

Academia generally accepts certain logical criteria for assessing meritocracy, such as having authored several high-impact articles as a principal investigator and actively engaging in research and teaching. The challenge lies in translating these expectations into a straightforward yet comprehensive mathematical algorithm, one that can also be optimized for different specific cases. Achieving this requires combining various metrics (Jauch & Glueck, 1975), but the necessary computer codes for processing these metrics are often unavailable. Consequently, university departments or academic competition committees often devise their own methods for evaluation and classification. However, without sufficient expertise in scientometrics, these officials may resort to improvised approaches that appear reasonable on the surface but are ultimately flawed. Such methods can lead to non-transparent outcomes that contradict the statutes of the universities. Evaluation criteria are considered "*transparent*" when they effectively and thoroughly highlight the genuine merit and qualifications of the individuals being assessed. In contrast, they are deemed "*non-transparent*" or "*biased*" when they only highlight selective aspects of the candidates' qualifications.

Several issues must be addressed. First, extracting relevant information from the vast array of publication and citation databases can be challenging. Second, it is critical to understand the exact meaning of proposed bibliometric metrics to avoid any misuse. Third, comparisons must be fair among individual researchers from different academic disciplines that might have varying publication and citation densities. Finally, creating an effective algorithm and corresponding computer code may require some mathematical, statistical, and programming skills. These and other relevant concerns are well articulated in the "*Leiden Manifesto for Research Metrics*" (Hicks et al., 2015) and in the "*Agreement on Reforming Research Assessment*" (CoARA, 2022).

Evaluation processes can suffer from significant shortcomings, often lacking transparency due to various factors. These include reliance on inappropriate and partial metrics, exclusively use of a single metric, and the application of non-standardized functions that overlook the nuances of different scientific disciplines. Inadequate assessments can also result from the improper use of readily available metrics made available by Elsevier SCOPUS (Schotten et al., 2017) and Clarivate Web of Science.

Pitfalls often emerge when metrics designed for one purpose are misapplied to another. A notable example is the uncritical use of the journal Impact Factor or CiteScore to gauge the importance of a published article. This practice is widely regarded as "*inappropriate*" (cf. CoARA, 2022; Hicks et al., 2015). The logical fallacy arises because journal impact factors and similar metrics assess the quality of an entire journal rather than that of individual research articles. These metrics were originally created to assist librarians in identifying which journals to subscribe to. Notably, even high-impact journals like Nature and Science can publish flawed or low-quality articles (Noorden, 2023), illustrating that "*using journal rank as an assessment tool is bad scientific practice*" (Brembs et al., 2013).

Indeed, only about 10% of Nobel Prize-winning research in physics is published in prestigious journals like Science and Nature (Bjørk, 2020). This indicates that the most innovative scientific research is not always found in the most esteemed journals. For recently

published articles that may not yet have accrued sufficient citations, it remains debatable whether the prestige of a journal could serve as the most reliable indicator of an article's significance.

Ultimately, the scientific community determines the relevance of a scientific publication, a process that may require time. Generally, scientific journals serve as platforms for showcasing scientific work. Therefore, publishing in a high-impact journal is desirable, as it is likely to attract attention and be viewed as relevant by a larger audience of scientists. Since such articles are more likely to receive public attention and citations, their relevance should not be double-counted by adopting metrics that consider also journal rankings. In this context, the San Francisco Declaration on Research Assessment (DORA) questioned using journal impact factors and similar metrics for assessing a scholar's scientific prestige. These metrics should not be used as "*measure of the quality of individual research articles, nor in hiring, promotion, or funding decisions*" (Alberts, 2013), though they may still be applicable for reasonably evaluating universities, departments, or research institutions as a whole.

Another common pitfall is the reliance on single citation metrics, such as the number of published articles or the H-index, mainly for convenience. Generally, using a single metric can be inappropriate and misleading. The number of articles alone is not a reliable measure; one researcher may publish several articles with modest scientific interest, while another may produce only a few but high-impact works. The H-index, defined as the number of articles cited at least H times, was proposed by Jorge E. Hirsch (2005) to balance the volume of published articles with the citations they receive from the scientific community. While the H-index is considered a more reliable indicator of excellence than article count alone (Bornmann & Hans-Dieter, 2018), it has significant limitations. Since 2010, when the H-index gained popularity, its correlation with scientific awards has decreased. This can be attributed to Goodhart's law, which states, "*when a measure becomes a target, it ceases to be a good measure*" (Strathern, 1997).

In fact, single metrics can be easily manipulated by changes in the behavior of the evaluated community. In the case of the H-index, the phenomenon of "*hyper-authorship*" has emerged, characterized by articles with numerous co-authors, sometimes exceeding 100 or even 1,000 (Nogrady, 2023). This practice artificially inflates citation counts and the H-index of each author, as co-authors tend to cite their own work. To mitigate the effects of hyper-authorship bias, Schreiber (2008) proposed a revised version of the H-index called the "*Hm-index*", which considers the number of authors involved in a publication.

Nonetheless, both the H-index and the Hm-index have limitations, as they do not account for an individual author's contribution to a work. While it is straightforward to credit a single author for a one-author article, determining the relative contribution of each author in a multi-authored work can be complex. Typically, the first author is the one who contributed the most, while other authors may have made lesser contributions and should not be regarded as principal investigators. Identifying a scientist's contribution to each published article is essential, especially when there are more than three authors. As a result, several journals now require a short paragraph detailing each author's specific contributions.

Generic evaluation criteria that fail to differentiate between first or principal investigators and other coauthors inevitably lead to distorted rankings when assessing and comparing the academic careers of a group of scientists. Equally problematic is the imposition of rigid time constraints, such as considering only works published within the last 5, 10, or 15 years. It is worth noting that Albert Einstein was awarded the Nobel Prize in 1921 for a work published 16 years earlier in 1905; Nobel laureates are typically awarded 12 to 33

years after their discoveries (Bjørk, 2019). Except in rare, highly specific cases, such criteria should be regarded as fundamentally "*non-transparent*" and, therefore, incompatible with the statutes of the universities when evaluating and ranking a group of scientists for academic promotion, such as the advance from associate to full professorship. They create an unjust framework, potentially equating scientists with vastly different levels of academic prestige and enabling secondary considerations to skew true ranking. For instance, a seasoned scientist with 30 years of experience and 120 publications—60 of which were authored as a principal investigator in the past decade—could be unfairly ranked on par with a younger scientist who has produced only 60 works, all as a mere coauthor within the same time frame. Such evaluation practices fail to acknowledge the genuine differences in contribution and expertise among scientists, rendering them "*non-transparent*", as schematized in Fig. 1.

Subtle challenges emerge when the aforementioned pitfalls are embedded within widely used evaluation systems, which could be misused to create improvised criteria for assessing and ranking the academic performance of a group of scientists, despite being designed for entirely different purposes. For instance, the Italian National Agency for the Evaluation of Universities and Research Institutes (ANVUR, https://www.anvur.it/en/homepage/, accessed on February 7, 2025) developed two such systems: the "*National Scientific Habilitation*" (ASN, "*Abilitazione Scientifica Nazionale*") for academics pursuing professorships, and the "*Research Quality Assessment*" (VQR, "*Valutazione della Qualitá della Ricerca*") for universities and research institutes. The metrics underpinning these systems are fundamentally ill-suited for comparing and ranking

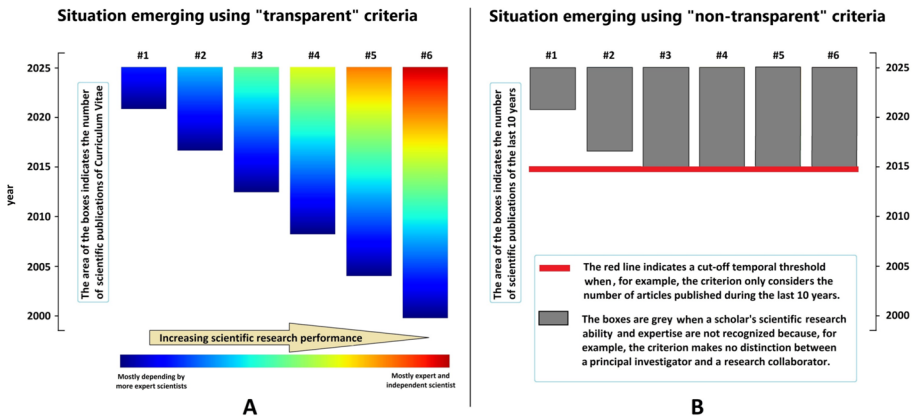

Fig. 1 Schematic comparison of academic rankings emerging from "*transparent*" and "*non-transparent*" evaluation criteria. (A) Illustration of the actual scientific ranking among six associate professors in a department, which emerges by adopting "*transparent*" evaluation criteria, where Associate Professor #6 stands out due to having the highest number of publications (represented by the larger box) and the highest level of individual prestige such as being the principal investigator of the published scientific works (indicated by the color gradient representing scientific leadership). (B) Illustration of the same scenario emerging under "*non-transparent*" criteria, such as the application of a sharp temporal cut-off (red line) that restricts the analysis to publications from the last ten years and the disregard of individual scientific leadership maturity as reflected by the uniform gray coloring of the boxes. Under these filtered conditions, Associate Professors #3, #4, #5 and #6 are artificially portrayed as being at the same scientific level, enabling secondary considerations to distort the true ranking highlighted in panel A and potentially favor Associate Professors #3, #4, and #5 over Associate Professor #6

the academic careers of a group of scientists like the faculty members of a department. This is due to their reliance on rigid time thresholds (i.e., considering only works and citations published during the last 5, 10 and 15 years) and their failure to account for varying levels of author contributions to the evaluated works. The misuse of ASN and VQR, and other similarly "*non-transparent*" evaluation criteria can significantly skew the rankings of faculty members and even foster forms of "*academic inbreeding*" such as "*parochialism*" and "*particularism*" by unfairly favoring junior assistants while indirectly discriminating against more experienced scientists who may lack political alignment within their departments; practices widely criticized as "*nefarious*" to the core functioning and mission of universities, as highlighted in international literature (Horta, 2022).

To accurately and transparently assess a researcher's scientific performance, standardizing and aggregating various quantitative and complementary bibliometric variables is crucial. Using multi-composite metrics based on complementary indicators diminishes the chance of manipulation, as illustrated by Goodhart's law. For example, the artificially inflated H-index values from hyper-authorship can be counterbalanced by the correspondingly low Hm-index values that such collaborative works often receive. Therefore, any fair metrics should necessarily combine the H-index and Hm-index to ensure that high individual scores are reflected in both indices.

Considering some of the above aspects, Ioannidis et al. (2016, 2019, 2020) proposed a standardized citation metric known as the "*C-score*". The C-score combines six bibliometric indicators: total citations, the H-index, the coauthorship-adjusted Hm-index, and citations of publications with three different authorship positions (single author, first author, and last author). These indicators are aggregated into a single metric. The work of Ioannidis et al. has garnered international attention, leading to the creation of a database of top-cited scientists with standardized citation indicators, known as the "*Stanford/Elsevier World Top 2% Most Influential Scientists*" list. The list is updated every year in two formats ("*Career-long Impact*" and "*Recent Year Impact*"); the latest update to date was published in August 2024 (Ioannidis, 2024). This database is utilized by academic institutions around the globe to recognize their most cited researchers. Furthermore, the Joint Research Centre of the European Commission of the European Union (Hollanders et al., 2018, p. 85) has also endorsed the C-score to identify researchers with the highest citation rates.

The C-score is widely recognized as a reliable assessment tool; however, it has its limitations. Generally, international global rankings may not be useful in local or small settings (cf. CoARA, 2022). For instance, the published list generated by the algorithm includes only 2% of the world's best scientists, which is rarely effective for ranking the faculty members within a department since only a few of them are likely to be included in such a list.

Additionally, the C-score does not account for other factors considered relevant in academia. For example, while imposing rigid time thresholds is unfair, it is useful to assess an academic's performance over time, as some scientists may no longer be actively engaged in research, and this should be reflected in their ratings. Furthermore, university departments often need to compare scholars from different academic disciplines, each characterized by varying bibliometric statistics. Evaluation metrics should also include works that are not typically indexed in citation databases, such as books, book chapters, patents, other scholarly contributions like research projects, and also works that are not written in English. It is also essential to consider the significant role of second authors in the development of research papers. Finally, aspects related to teaching experience and other university missions also contribute to a professor's overall performance and should be considered in the evaluation metrics.

Building on these considerations, this paper introduces a new metric that integrates the C-score algorithm. This new methodology develops a composite index that accounts for additional factors, such as publication volume, co-authorship, different academic fields, and career progression over time. This proposed measure is further enhanced by factors typically considered for academic appointments, including teaching experience and the ability to secure funding for research and other academic services. The accompanying computer code utilizes SCOPUS bibliometric files, which are readily available online and can easily be adjusted, if necessary, to generate an Excel file containing the desired RT-score rankings of a group of scientists. The algorithm also relies on several free parameters, allowing users to tailor it to specific contexts and requirements.

This research, along with the accompanying code, was motivated by the desire to support university departments in evaluating the research performance of their faculty members, as well as in other similar situations where effective evaluation algorithms are unavailable.

A comprehensive standardized research-teaching (RT) score

The proposed "research" module of the RT-score algorithm creates a composite and standardized research indicator by building upon the C-score algorithm (Ioannidis et al., 2019, 2020). The modifications and applications of the original C-score formula aim to enhance the algorithm's effectiveness in evaluating the scientific performance of a group of researchers in a typical academic context. It is assumed that a reliable R-score should integrate and weigh at least the following factors:

- The total number of scientific publications;
- The number of citations for those scientific publications;
- The total number of co-authors;
- The individual's personal contribution to the publications;
- A higher score assigned to a select group of high-impact papers;
- The time progression of the above indices, with greater weight assigned to more recently published articles;
- The indices should be appropriately scaled when comparing scientists from different academic fields.

Standardized citation metrics: basic formula

The proposed R-score equation includes seven bibliometric indicators that balance impact (number of citations) and productivity (number of publications), while also considering the number of authors and the researcher's position in the authorship list. The composite research score, "R", for researcher "i" (where i = 1, ..., N) uses the following indicators:

1. NA—the number of articles published and indexed in SCOPUS by the author, which can be adjusted to include any other work not indexed in SCOPUS.
2. H (Hirsch index)—the number of articles with at least H citations (Hirsch, 2005).
3. Hm (Schreiber index)—it is similar to the H-index but accounts for the number of authors by dividing citations by the total authorship (Schreiber, 2008).
4. NC—total citations received by all articles.

5. NCS—citations of the author's articles as a single author.
6. NCSF—citations of the author's articles either as a single author or first author (NCSF ≥ NCS).
7. NCSFSL—citations of the author's articles as a single author, first author, or second/last author.

The indices mentioned above need to be standardized and normalized relative to the group of N scientists. It is proposed to normalize the number of publications (NA) against the maximum value of NA observed among the group. The resulting ratio will then be processed using the square root function. This approach is recommended because the number of publications by the same author tends to increase over time but often overlaps, as authors typically focus on specific topics that evolve gradually.

However, the provided MATLAB code includes optional subroutines for those who prefer to use different approaches. One option employs a linear function, represented as $(NA_i + 1)/max_{j=1,...,N}(NA_j + 1)$, indicating that all works are very innovative. Another option uses a logarithmic function, expressed as $\log(NA_i + 1)/max_{j=1,...,N} \log(NA_j + 1)$, suggesting a strong overlap among the works.

As proposed by Ioannidis et al. (2019, 2020), the indicators H, Hm, NC, NCS, NCSF, and NCSFSL are logarithmically normalized based on the maximum value of each indicator within the ensemble of N scientists. This normalization is performed because citation distributions can often be approximated by some type of exponential function (Vieira & Gomes, 2010). Finally, each scientist is assigned the resulting R-score:

$$
\begin{aligned}
R_i = \Bigg[& \frac{r_1 \cdot \sqrt{NA_i + 1}}{max_{j=1,...,N} \sqrt{NA_j + 1}} + \frac{r_2 \cdot \log(H_i + 1)}{max_{j=1,...,N} \log(H_j + 1)} \\
& + \frac{r_3 \cdot \log(Hm_i + 1)}{max_{j=1,...,N} \log(Hm_j + 1)} + \\
& \frac{r_4 \cdot \log(NC_i + 1)}{max_{j=1,...,N} \log(NC_j + 1)} + \frac{r_5 \cdot \log(NCS_i + 1)}{max_{j=1,...,N} \log(NCS_j + 1)} \\
& + \frac{r_6 \cdot \log(NCSF_i + 1)}{max_{j=1,...,N} \log(NCSF_j + 1)} + \\
& \frac{r_7 \cdot \log(NCSFSL_i + 1)}{max_{j=1,...,N} \log(NCSFSL_j + 1)} \Bigg] \cdot \frac{100}{\sum_{i=1}^{7} r_i}.
\end{aligned}
\tag{1}
$$

In this context, the function $max_{j=1,...,N}$ identifies the maximum index value among the N scientists. Equation 1 assigns a score R_i to each researcher, ranging from 0 to 100, and these scores can be ranked in descending order. Users can modify the metrics selecting different weights $(r_1, ..., r_7)$; the default weights are set to $r_{1,...,7} = 1$. To transform Eq. 1 into the C-score proposed by Ioannidis et al. (2019, 2020), it is necessary to set $r_1 = 0$ and $r_{2,...,7} = 1$, while ignoring the second author for NCSFSL.

The variable NA can be adjusted by editing the SCOPUS file "name.scopus.cvs," allowing for the inclusion of additional publications not indexed in SCOPUS, like books, research proposals and patents, with assigned weights. For example, to account for a book that is considered equivalent to five articles, users must replicate the line five times in the file, setting the citation number to "0." Instructions for handling this kind of files are in the Appendix.

Mathematically, NC ≥ NCSFSL ≥ NCSF ≥ NCS, indicating an increasing weight for papers based on authorship: sole authors receive the most weight because these papers will be considered in all four metrics, followed by first authors, then second or last authors, and finally all the other authors. Last authors should merit additional recognition because they may have supervised the research group, particularly when the authors' names are not listed alphabetically. Although corresponding authors might also merit additional recognition, the algorithm does not allocate extra points to this category, as the SCOPUS file does not specify the corresponding author.

The SCOPUS files do not indicate whether the last author oversaw the study group. Consequently, the proposed MATLAB code does not differentiate between these two scenarios by default; the C-score does the same (cf.: Ioannidis et al., 2019; 2020). However, the code includes an optional subroutine that allows the last author to be considered only if their last name is not in alphabetical order relative to the second-to-last author. This option should be used with caution, as an author whose name begins with a letter like "Z" is likely to be the last author in both cases, which could lead to confusion. Therefore, by default, the code always allocates extra points to the last authors (as implemented in the C-score) and associates them with the second authors. Single and first authors receive the highest score, followed by second and last authors. The remaining authors receive only the basic score for their contribution to the work. This approach ensures that the three primary authors of each paper always receive some additional score. Scientists who typically publish in large groups can enhance their R-score by publishing numerous high-impact papers and/or assuming the role of principal investigators.

Assuming differing article number and citation densities for ranking a partially heterogeneous group of scholars

Comparing the scientific bibliographies of scholars from different disciplines can be essential in academia. Various scientific fields often exhibit statistical differences in metrics such as the number of articles, H-indexes, and citation densities. This challenge can be addressed by applying appropriate scaling factors for each field. In Italy, one potential approach is to use the ASN thresholds for each scientific discipline (known as "*Settore Scientifico-Disciplinare*" or SSD) as benchmarks. This section discusses how this can be accomplished.

For each SSD, ASN offers three thresholds (https://www.anvur.it/attivita/asn/asn-2018-2020/indicatori-e-relative-soglie/, accessed on Feb/07/2025):

1. The minimum number of papers published in the last 10 years;
2. The minimum H-index in the last 15 years;
3. The minimum number of citations in the last 15 years.

These thresholds can be utilized to generate scaling factors for each SSD, denoted as α_i, β_i and γ_i. This is achieved by calculating the ratio of the maximum threshold value observed among the N scholars in the group to the threshold value of the i^{th} scholar, represented as "Tr_i":

$$\text{scaling}_i = \frac{\max_{j=1,\ldots,N}(Tr_j)}{Tr_i}. \tag{2}$$

Table 1 provides an example derived from the discipline of "Earth Sciences", which in Italy is divided into 12 SSD. The resulting scaling coefficients α_i, β_i and γ_i for $i = 1, ..., N$ can be employed to adjust the Eq. 1 as follows:

$$
\begin{aligned}
{}^{\text{adj}}\text{R}_\text{i} = \Bigg[& \frac{r_1 \cdot \sqrt{\alpha_\text{i} \cdot \text{NA}_\text{i} + 1}}{\max_{\text{j}=1,\ldots,\text{N}} \sqrt{\alpha_\text{j} \cdot \text{NA}_\text{j} + 1}} \\
& + \frac{r_2 \cdot \log(\beta_\text{i} \cdot \text{H}_\text{i} + 1)}{\max_{\text{j}=1,\ldots,\text{N}} \log(\beta_\text{j} \cdot \text{H}_\text{j} + 1)} \\
& + \frac{r_3 \cdot \log(\beta_\text{i} \cdot \text{Hm}_\text{i} + 1)}{\max_{\text{j}=1,\ldots,\text{N}} \log(\beta_\text{j} \cdot \text{Hm}_\text{j} + 1)} + \\
& \frac{r_4 \cdot \log(\gamma_\text{i} \cdot \text{NC}_\text{i} + 1)}{\max_{\text{j}=1,\ldots,\text{N}} \log(\gamma_\text{j} \cdot \text{NC}_\text{j} + 1)} \\
& + \frac{r_5 \cdot \log(\gamma_\text{i} \cdot \text{NCS}_\text{i} + 1)}{\max_{\text{j}=1,\ldots,\text{N}} \log(\gamma_\text{j} \cdot \text{NCS}_\text{j} + 1)} \\
& + \frac{r_6 \cdot \log(\gamma_\text{i} \cdot \text{NCSF}_\text{i} + 1)}{\max_{\text{j}=1,\ldots,\text{N}} \log(\gamma_\text{j} \cdot \text{NCSF}_\text{j} + 1)} + \\
& \frac{r_7 \cdot \log(\gamma_\text{i} \cdot \text{NCSFSL}_\text{i} + 1)}{\max_{\text{j}=1,\ldots,\text{N}} \log(\gamma_\text{j} \cdot \text{NCSFSL}_\text{j} + 1)} \Bigg] \cdot \frac{100}{\sum_{i=1}^{7} r_i}.
\end{aligned}
\tag{3}
$$

The α_i coefficients are applied to scale the number of articles, the β_i coefficients scale the H-index and the modified H-index (Hm), while the γ_i coefficients scale the other four metrics related to citations. If all N scholars belong to the same SSD, such as when competing for a specific academic position, it suffices to set $\alpha_i = \beta_i = \gamma_i = 1$ for each candidate

Table 1 The highest ASN thresholds (tr) for the position of Commissioner ("*Commissario*") for the 12 academic disciplines (SSD) in the field of Earth Sciences

SC	SSD	Tr. N. articles	Tr. H-index	Tr. N. citations	Scaling factors		
		(10 years)	(15 years)	(15 years)	α	β	γ
04/A2	GEO01 (GEOS-01/A)	26	15	570	1.19	1.00	1.16
04/A2	GEO02 (GEOS-01/B)	26	15	570	1.19	1.00	1.16
04/A2	GEO03 (GEOS-01/C)	26	15	570	1.19	1.00	1.16
04/A3	GEO04 (GEOS-02/A)	18	11	216	1.72	1.36	3.06
04/A3	GEO05 (GEOS-02/B)	18	11	216	1.72	1.36	3.06
04/A1	GEO06 (GEOS-03/A)	31	15	662	1.00	1.00	1.00
04/A1	GEO07 (GEOS-03/B)	31	15	662	1.00	1.00	1.00
04/A1	GEO08 (GEOS-03/C)	31	15	662	1.00	1.00	1.00
04/A1	GEO09 (GEOS-03/D)	31	15	662	1.00	1.00	1.00
04/A4	GEO10 (GEOS-04/A)	25	11	360	1.24	1.36	1.84
04/A4-Geo11	GEO11 (GEOS-04/B)	17	10	293	1.82	1.5	2.26
04/A4	GEO12 (GEOS-04/C)	25	11	360	1.24	1.36	1.84
maximum	"MAX"	31	15	662			

The scaling factors α, β and γ are evaluated using Eq. 2. (Thresholds for each SSD are taken from https://www.anvur.it/attivita/asn/asn-2018-2020/indicatori-e-relative-soglie/, accessed on Feb/07/2025)

$i = 1, ..., N$. The adjusted scores ${}^{adj}R_i$ are then utilized to create the desired ranking among the N academics.

The provided MATLAB code computes individual scores for both the actual and scaled metrics but only ranks the scaled metrics. These scaling factors ensure that, for instance, if one scholar has published 100 articles and another has published 200, with their corresponding thresholds also being 100 and 200 respectively, both scholars are assessed as having performed similarly based on such metric. This requirement aligns with point 6 of the Leiden Manifesto (Hicks et al., 2015).

Combining the *Curriculum Vitae* and time-progression standardized citation metrics

Institutions or departments often require more comprehensive and standardized citation metrics than those represented by by Eq. 3. These metrics should evaluate a selection of the best-published papers while also incorporating time progression metrics. The latter is essential because scientists no longer active in research may not qualify for promotion, yet they can still be considered for awards (such as the Nobel Prize) for their past work.

An effective algorithm should gradually increase the weight of recent publications while also prioritizing a limited set of a researcher's most significant works. According to Italian law, the minimum number of publications for a formal evaluation in a professorship competition should not be less than 10. In actual full professorship competitions, the number of publications selected for detailed evaluation typically ranges from 15 to 30.

A comprehensive index can be constructed by evaluating the measure represented by Eq. 3 in several specific scenarios, averaging the derived coefficients ${}^{adj}C_i$ as necessary, and ranking the list of N scores representing each author. In this section, this paper proposes an equation that combines four metrics:

1. ${}^{all.adj}R_i$—standardized citation metrics relative to the entire curriculum vitae of each of the N researchers;
2. ${}^{30a.adj}R_i$—standardized citation metrics relative to the 30 most cited papers from the curriculum (the number of papers can be adjusted as desired);
3. ${}^{15y.adj}R_i$—standardized citation metrics relative to papers published in the last 15 years;
4. ${}^{5y.adj}R_i$—standardized citation metrics relative to papers published in the last 5 years.

The proposed comprehensive, standardized citation metric is obtained by averaging the above four measures with weights $w = (w_1, w_2, w_3, w_4)$ as

$$ {}^{tot}R_i = \frac{w_1 \cdot {}^{all.adj}R_i + w_2 \cdot {}^{30a.adj}R_i + w_3 \cdot {}^{15y.adj}R_i + w_4 \cdot {}^{5y.adj}R_i}{w_1 + w_2 + w_3 + w_4}. \tag{4} $$

The default weights are $w_1 = w_2 = w_3 = w_4 = 1$, and the MATLAB code allows users to modify them as needed. The default values (30 articles, 15 years, and 5 years) can also be adjusted according to user requirements. Ultimately, the total score ${}^{tot}R_i$ is used to determine the desired research ranking among the N scientists.

Due to its mathematical structure, Eq. 4 employs a mixed and nested framework, integrating two measures of the curriculum vitae (${}^{all.adj}R_i$ and ${}^{30a.adj}R_i$) with two CV time progression measures (${}^{15y.adj}R_i$ and ${}^{5y.adj}R_i$). This approach allows Eq. 4 to consider an academic's entire career while giving greater weight to an author's most prestigious and recent publications.

Scoring teaching expertise, the ability to acquire funding and other factors

The scientific research performance of a scientist can be effectively evaluated using Eq. 4. The second most important component of the evaluation process in universities is teaching experience. Other factors, such as the ability to obtain funding, departmental service, etc., contribute to the evaluation process to a lesser extent. The following subsections suggest how these factors could be taken into account as well.

Teaching experience

The simplest way to assess a professor's teaching experience is to count the number of credit hours (referred to as "Nh_i") he or she has taught at a university. In Italy, 1 credit hours usually corresponds to 8 h of class teaching. For a group of N academics, the relative teaching score T_i for i = 1, ..., N can be defined as:

$$T_i = 100 \frac{\sqrt{Nh_i + 1}}{\max_{j=1,...,N} \sqrt{Nh_j + 1}}. \tag{5}$$

In Eq. 5, a modulation based on the square root function is preferred by default as a compromise between linear and logarithmic modulation, which are provided as options in the MATLAB code. The reason for this is that teaching experience increases over time, but only gradually, as courses are often repeated with moderate changes.

It should be noted that the algorithm suggests the use of the number of teaching hours because it is a reasonable and objective quantitative measure. However, the number of teaching hours can be replaced with any other index deemed appropriate. For example, the user may develop a metric that also takes into account the number of students, the number of theses and dissertations supervised, using appropriate scaling factors. Such user-defined indices may also attempt to assess, for example, the qualitative aspects of teaching, if available, or add other academic activities including also those that pertain the "*Third Mission of the University*", i.e. the dissemination of knowledge outside the academia for social, cultural and economic development (Compagnucci & Spigarelli, 2020). Thus, the number of teaching hours could be supplemented with other forms of teaching activity, such as presentations at conferences, seminars, mass-media activities, etc.

It may also be possible to implement this voice with an additional qualitative estimate of the research impact or the public interest in the research (such as number of readers, blog mentions, news mentions, etc.), which are other very important voices to properly assess the scientific prestige of an academic (cf.: CoARA, 2022; Hicks et al., 2015). In this respect, an increasing number of scientific journals are providing information on the public attention received by their published articles through services such as Altmetric (https://www.altmetric.com/), Plumx (https://plu.mx/), and others.

Then, the chosen integrated metric, labeled ${}^{tot}T_i$ score, can be used in the T-score algorithm instead of the number of traditional teaching hours in class. To keep it simple, one can add the number of teaching hours in class and the number of scientific presentations including those outside academic environments.

Ability to acquire funds

The ability to secure funding is also recognized as a valuable skill in academia, yet there is a need for effective methodologies to assess this skill and integrate it into the RT-score. The problem with properly assessing this ability is that, unlike research and teaching, fundraising is not a core mission of universities. Funding serves as an input to research rather than a goal. While funding can lead to higher-quality research (Thelwall, 2023), notable scientists like Gregor Mendel, Michael Faraday, Albert Einstein, Max Planck and many others often only needed basic financial support from their institutions, and other prestigious scientists (e.g.: Charles Darwin, Henry Cavendish, Stephen Wolfram, James Lovelock, James Hutton, and many others) self-funded much of their research. Ultimately, the impact of research results, not the amount of funding awarded to a scientist by external sources, determines scientific quality.

Scientific literature questions whether fundraising ability is relevant for evaluating scientists. For instance, Gillett (1991) critically examines the strategy of judging the quality of scientific research based on the level of funding it attracts. He argues that an index such as per capita research income, which relies on grant-giver peer review, provides an unsatisfactory measure of scientific performance because it fails to meet a fundamental requirement for performance indicators: the relationship between "*outputs*" and "*inputs*". Gillett contends this method has low validity and is often confounded by various extraneous factors unrelated to research performance. These may include political and economic influences and conflicts of interest, which appear to be particularly significant in fields like biomedical and climate science (Wojick & Michaels, 2015). He further emphasizes "*the inappropriateness of using an input measure as a performance indicator is highlighted by the fact that a department*" (or a scientist) "*with many grants would automatically be awarded a high score even if it produced no research at all*". Laudel (2005) states: "*the study demonstrates that success in obtaining external funding is only partly related to the quality of researchers and their proposals. Therefore, the validity of a straightforward counting of external funding must be assumed to be low*". He adds: "*Availability of funding depends on the actual topic and might therefore vary even within fields. In this context, the basic/applied character of research must also be taken into account. The more applied a researcher's work is, the more potential funding sources outside academia (industry, government, etc.) exist*."

Thus, the primary challenge is to assess funding acquisition abilities without penalizing theoretical scientists whose research may not require external funding. Typically, only experimental and applied research needs substantial financial support. To address this, the following criteria are proposed:

- Research proposals undergo peer review, like formal research publications, and, therefore, they can be classified as non-indexed written works (like books, book chapters and patents) not listed in SCOPUS. These should be counted in the number of articles (NA) in Eq. 3, applying scaling factors based on their significance. For instance, a modest project may count as one publication, while a larger one could equate to five or ten publications. A practical approach might be to count twice the number of publications produced by the funded project. This approach boosts the NA weight of funded researchers. The operation can be done by editing the SCOPUS publication file as detailed in Appendix.

- The results of funded projects should be published as regular scientific articles, whose scientific relevance is already evaluated using the existing R-scores (Eqs. 1, 3 and 4). Consequently, no additional measures are considered necessary in this respect because "*double counting*"—counting both the grant and the resultant publications—must be avoided, as the research output relies on its financial input (Johnes & Johnes, 1995).
- The additional administrative activities of the Principal Investigator can only be assessed qualitatively and combined with other faculty responsibilities. Thus, a direct quantitative assessment is not possible, and these activities generally represent a small fraction of an academic's overall rating. Some metrics may be developed and added in the T-score with appropriate scaling.
- Academic promotions should primarily focus on research and teaching performance, rather than funding amounts. However, financial support may lead to a personal economic benefit, such as a salary increase during the funded project.

The suggestions presented aim to clarify that funding should not be overemphasized as a marker of academic performance (Gillett, 1991; Laudel, 2005). An academic's ability to secure additional funding—beyond standard university research funding—should result in a moderate increase in his/her scientific performance score. This increase could objectively be reflected in an enhancement of the NA index, which complements the basic R-score assigned to published works. In general, funding should be perceived as a resource for specific research initiatives, not as an indicator of research quality by itself, and a researcher's academic standing depends only on the quality of their scientific output, regardless of the source of funding. Nonetheless, attracting additional funding should provide a legitimate economic advantage to those who bring financial resources into their university departments. For instance, many U.S. institutions compensate professors for only nine months of the year, and grants can increase their annual salary by up to 33% during the remaining months.

It is crucial to acknowledge that the work of theorists who rely on reasoning and theoretical models should not be undervalued or penalized solely due to the absence of unneeded formal project grants. Such practices risk breaching academic principles enshrined in university statutes, which safeguard research freedom and equal opportunities, stipulate that the quality of research should be assessed according to specific criteria and indicators decoupled from purely economic logic, and prohibit any forms of discrimination, including the indirect ones.

It is well documented in the history of science that significant theoretical contributions can be made even in the absence of external funding. Furthermore, an overemphasis on the ability to secure funds overlooks the fact that very costly projects are often conducted to produce large experimental databases that are published and then analyzed by theorists around the world. Notably, many Nobel laureates (including Albert Einstein) made significant advancements by interpreting experimental results (e.g. the Michelson-Morley experiment, 1887) obtained by other groups, which were those that required some funding for the experimental equipment. These theorists resolved important physical problems with minimal financial resources. Indeed, academics are expected to contribute to society by advancing knowledge, rather than solely focusing on securing some funding for their departments (cf.: CoARA, 2022; Hicks et al., 2015).

Creation of a comprehensive standardized research-teaching (RT) score

Finally, the ${}^{\text{tot}}\text{T}_\text{i}$ score can be combined with the ${}^{\text{tot}}\text{R}_\text{i}$ score to produce the total Research-Teaching (RT) total score (${}^{\text{tot}}\text{RT}_\text{i}$) by using a chosen weighting factor RT_{rw} as

$$ {}^{\text{tot}}\text{RT}_\text{i} = \text{RT}_\text{rw} \cdot {}^{\text{tot}}\text{R}_\text{i} + (1 - \text{RT}_\text{rw}) \cdot {}^{\text{tot}}\text{T}_\text{i}. \tag{6} $$

For example, if RT_{rw} is set to 0.70, then 70% of the total score will be attributed to research and 30% to teaching. If RT_{rw} is set to 1.00, teaching will be excluded from the evaluation.

Jauch and Glueck (1975) noted that several additional factors could be used to assess the performance of a university lecturer. However, the ${}^{\text{tot}}\text{RT}_\text{i}$ score provided in Eq. 6 could serve as a sufficient basis for an evaluation that could be considered "*transparent*" in most practical scenarios because the proposed metrics could be sufficiently comprehensive, especially if the NA variable (Eq. 3) is adjusted to include funded research proposals, and if the teaching component encompasses scientific presentations, seminars, and other academic activities related to the University's Third Mission.

Since any evaluation algorithm should not be overly simplistic yet remain straightforward (cf.: CoARA, 2022; Hicks et al., 2015), any additional criteria could be incorporated into the T-score based on specific needs. Alternatively, these criteria can be ignored as a first approximation, as they tend only to differentiate between academics with very similar RT scores and can only be derived from a specific qualitative analysis on a case-by-case basis.

A toy application involving 5 outstanding scientists

Table 2 presents a toy application of the RT-score algorithm to five recent Nobel Prize laureates in physics (Pierre Agostini, Alain Aspect, Geoffrey Hinton, John J. Hopfield, and Giorgio Parisi) to illustrate how the proposed approach works. For simplicity, the same default variables are assumed and that the academic field parameters are $\alpha = \beta = \gamma = 1$, the teaching score is T-score = 100, and $RT_{rw} = 0.70$ for all five scholars. Table 2A shows the total scores from the "RT_ranking.xlsx" output file generated by the RT-score algorithm, while Table 2B provides a breakdown of the calculations across the proposed four independent tests: (1) full SCOPUS curriculum vitae (CV), (2) top 30 articles (Top30), (3) articles from the last 15 years (2010–2025) (Last15y), and (4) articles from the last 5 years (2020–2025) (Last5y).

The rankings of these scientists vary depending on the method of calculation. For example, Hopfield ranks third in the overall R-score but first in the Top30 test, mostly because of his highly cited 30-top works written as single author. In contrast, Hinton ranks lower in the Top30 due to his high number of co-authors (83) compared to Hopfield (32), although he has an excellent overall R-score with his top 30 articles receiving 109,771 citations. High number of coauthors leads to a lower Hm score. Hopfield has published only two papers in the last five years, suggesting less recent research activity, which penalizes him in the Last5y test. Parisi, with 723 published articles, ranks second because his papers have fewer citations than Hinton in most tests. Agostini is penalized because he has exclusively written his works with other authors. The proposed toy application illustrates how the RT-score

Table 2 Application of the R-score: an example involving five recent Nobel Laureates in physics. [A] Summary of the R-score rankings for each of the 4 tests, and for the total R-score (Eq. 4). [B] Detailed statistics for each voice and the four tests. The SCOPUS files were downloaded on Feb/04/2025

[A]	Name	CV	Top30	Last15y	Last5y	R-score (0-100)	T-score	RT-score (0-100) (ranked)
Total	Hinton	93	87	96	94	92.50	100	94.75
R-Score	Parisi	92	59	76	72	74.25	100	82.33
	Hopfield	86	95	51	27	64.75	100	75.33
	Aspect	73	55	51	49	57.00	100	69.90
	Agostini	61	53	48	33	48.75	100	64.13

[B]	Name	N. Coa	N. Art	H	Hm	N.C	N.C.S	N.C.S.F	N.C.S.F.S.L	R-score
CV	Hinton	268	268	113	62	385741	9256	55404	379887	93
	Parisi	425	723	104	70	55815	7830	13302	40054	92
	Hopfield	106	189	74	51	49494	26608	36060	48809	86
	Aspect	423	226	54	22	18680	614	8529	13879	73
	Agostini	379	227	54	18	17295	0	2988	7156	61
Top30	Hopfield	32	30	21	13	19906	15236	15348	19690	95
	Hinton	83	30	23	7	109771	23	8780	108543	87
	Parisi	76	30	12	5	807	0	180	441	59
	Aspect	240	30	16	3	2994	0	2	1422	55
	Agostini	104	30	13	3	797	0	85	89	53
Last15y	Hinton	195	109	67	27	265578	1965	12564	263669	96
	Parisi	219	206	41	20	7525	29	1115	2720	76
	Hopfield	4	12	10	8	538	105	105	538	51
	Aspect	249	62	22	6	1520	20	26	87	51
	Agostini	130	67	22	5	3768	0	93	97	48

Table 2 (continued)

[B]	Name	N. Coa	N. Art	H	Hm	N.C	N.C.S	N.C.S.F	N.C.S.F.S.L	R-score
Last5y	Hinton	101	21	17	5	9760	23	23	9359	94
	Parisi	107	35	12	4	938	0	46	88	72
	Aspect	28	7	5	2	58	6	6	6	49
	Agostini	41	22	5	1	104	0	0	0	33
	Hopfield	3	2	2	1	55	0	0	55	27

metrics can be used to evaluate and rank scientists using their Scopus bibliometric data. However, all five scientists are not only exceptional in bibliometric terms, but also in terms of their significant contributions to science, which requires a thorough examination of their published results, which is something that also needs to be taken into account for a comprehensive evaluation (cf.: CoARA, 2022; Hicks et al., 2015).

Overall, the RT-score combines different and complementary tests to provide a bibliometric comparison between a group of scientists. Users can customize the default parameters, and the output file "RT_ranking.xlsx" provides metrics that can be interpreted according to specific contexts.

Conclusion

This paper introduces a novel algorithm, the RT-score, specifically developed to evaluate and rank the research and teaching performance of faculty members in university departments, academic competitions, and similar settings. The proposed algorithm can be regarded as sufficiently "*transparent*" because it effectively and comprehensively highlights the most typical qualifications of a scholar required in academia according to the scientometric literature. It is both straightforward and avoids excessive simplifications. Evaluators can readily adjust the code by setting appropriate values for its free parameters, while those being assessed can easily verify the data and analysis. These attributes are essential for ensuring academic fairness, equal opportunities, and transparency (cf.: CoARA, 2022; Hicks et al., 2015), while countering unethical practices such as the misuse of hyper-authorship and the implementation of biased criteria. Such criteria often lead to indirect discrimination through mechanisms like imposing rigid temporal thresholds, overlooking individual contributions to published works, and disproportionately emphasizing the ability to secure funding.

The RT-score algorithm primarily relies on available bibliometric information, using SCOPUS as the default database. To prevent misleading results, users must ensure that the SCOPUS files they download are complete and do not include papers not authored by the evaluated scientists. Occasionally, an author may mistakenly have multiple SCOPUS IDs, and the publication list may include works by synonymous authors. Users can verify or create publication files in SCOPUS format using the authors' submitted Curriculum Vitae.

To ensure completeness, it is also recommend supplementing the SCOPUS file with additional works, such as non-indexed articles, books, research projects, patents, and others that typically do not appear in SCOPUS. This may include works that are not published in English, which can be particularly significant in certain local contexts (cf. Hicks et al., 2015). For disciplines where SCOPUS is less relevant, such as the humanities or for scholars who do not publish in English-language journals, the proposed methodology may not perform well or be applicable because the SCOPUS archive may inadequately represent a professor's research activity.

The proposed RT-score algorithm builds upon the C-score algorithm of standardized citation metrics developed by Ioannidis et al. (2016, 2019, 2020). The C-score algorithm is well-known through the "*Stanford/Elsevier's World Top 2% Most Influential Scientists*" list and has been endorsed by the Joint Research Centre of the European Commission as a reliable metric for identifying highly cited researchers (Hollanders et al., 2018, p. 85). Despite some limitations (Philip, 2023), the C-score algorithm should provide a sufficiently robust foundation for ranking a group of scientists based on their publication and citation counts.

The RT-score algorithm introduces additional components to the C-score algorithm, enhancing its efficiency in classifying the scientific performance of a small group of scientists, such as faculty staff in a department or participants in an academic competition. The equation combines seven indicators: the total number of published articles, the H-index, the co-authorship-adjusted Hm-index, the total number of citations, and citations from articles in three distinct authorship positions (single author, first author, and second or last author). To ensure accurate rankings across different fields, the code includes three normalization factors that account for variations in the number of articles, H-index, and citation density across disciplines (Eq. 3). Additionally, the RT-score algorithm adopts a mixed/nested structure that incorporates two CV measures along with two temporal measures, giving more weight to the most cited and most recent publications (Eq. 4). Finally, a simple algorithm combines the publication score with the teaching score (Eq. 6), which can also include other relevant variables commonly used in academia. Considering both quantitative and qualitative factors the RT-score algorithm aligns with the key recommendations proposed by DORA, the "Leiden Manifesto for Research Metrics" (Hicks et al., 2015), and the "Agreement on Reforming Research Assessments" (CoARA, 2022).

Implementing evaluation criteria that are as "*transparent*" as possible is crucial, especially in contexts where legislative constraints and limited economic resources can intensify competition within a department. In such scenarios, there is a risk of indirect discrimination through the adoption of "*non-transparent*" criteria. Non-transparent criteria, as those exemplified in Fig. 1, can be easily exploited to manipulate the preliminary selection of the academic fields for opening full professorship competitions, as well as to influence the competition process itself (cf. Gallina and Gallo, 2020). For instance, in Italy, movements such as the "*Movimento Carlo Ferraro Dignitá Docenza*" (https://groups.google.com/g/docentipreoccupati/c/_-mmi0I8eG0, accessed Feb/07/2025), advocate for legislative reforms aligned with the tenure-track system to address such issues more equitably. This movement proposes that a tenured Associate Professor, after 25 years in academia (including at least 8 years in the role of Associate Professor) and having obtained ASN qualifications for Full Professor, should be entitled to an evaluation that allows him to advance to Full Professor.

An additional notable aspect of this study is the inclusion of MATLAB code that implements the proposed algorithm. The code is designed to be accessible and straightforward, allowing for immediate use or easy modification to meet specific needs. The Appendix provides detailed instructions for using the code.

It could be contended that the mathematical problem is more intricate than what is represented here, that no bibliometric index is without flaws, and that the development of more advanced mathematical methods and indices may be necessary. Nevertheless, in the absence of clearly defined evaluation frameworks, there is a significant risk of resorting to improvised algorithms that are mathematically unsound, poorly standardized, overly simplistic, and ultimately "*non-transparent*". As such, it is imperative for these proposals to be thoroughly developed and debated within the scientometric literature; otherwise, there is a danger that they will fall to the discretion of an evaluation committee, potentially leading to inconsistent and inadequate assessments, which may even lead to tensions among faculty members within a department and encourage the practice of "*academic inbreeding*" (Horta, 2022). This work aims to promote fairer evaluation practices in academia.

Appendix: How to use the provided MATLAB code " RT_method.m"

The material in the Supplement provides the MATLAB (R2024b) code, labeled "RT_method.m", along with a sample of SCOPUS "_.cvs" files from the five scientists discissed in Section 4. The code can be executed also in MATLAB online (available at https://matlab.mathworks.com/, accessed on Feb/07/2025) by creating an account and logging in.

The "RT_method.m" code is user-friendly and can be easily edited, as detailed in the included ReadMeFirst file. The default settings are outlined below.

1. For each of the N researchers, users must download bibliometric CSV files from SCOPUS (https://www.scopus.com/search/form.uri#author, accessed on Feb/07/2025) by using the "export all" function. The CSV files should only include three pieces of information: "Author(s)," "Year," and "Citation count." Users are responsible for manually deselecting any unnecessary fields. The CSV files can be renamed with the researcher's name and/or SCOPUS ID (e.g., "name.scopus.csv") and should be saved in the same folder as the MATLAB code.
2. If users need to manually edit the "name.scopus.csv" file to correct any information or to add works that are not indexed in SCOPUS (such as books, book chapters, research proposals, patents, etc.), they should not use MS Excel to open, edit, and save the file. Doing so may alter the cell formatting of the CSV file, which can lead to complications when the RT_method.m code attempts to read it. Instead, the file should be opened with a standard text editor, like Notepad (available with Windows 10 or 11) or Notepad++ (available at https://notepad-plus-plus.org/, accessed on Feb/07/2025). A new line should be added for each additional item.

   For example, if a book "Book1" values 3 points and a written research proposal values 2 points, the user must add five lines in the file "name.scopus.csv" in the following format:

   "nameScholar; book1 1/3", " ", " ", " year", " 0"
   " nameScholar; book1 2/3", " ", " ", "year", "0"
   "nameScholar; book1 3/3", " ", " ", "year", "0"
   "nameScholar; proposal1 1/2", " ", " ", "year", "0".
   "nameScholar; proposal1 2/2", " ", " ", "year", "0"

   In this format, "nameScholar" refers to the scholar's name as it appears in the SCOPUS file (e.g., Schmidt A.), and "year" indicates the specific year (e.g., 2022) when the work was produced.
3. To proceed, one needs to open the file named "RT_method.m" in the MATLAB editor. The first section of the code requires user editing and consists of seven variables along with the output file name. Here's a breakdown of what needs to be modified:

   (a) Year—the default value is 2025, which indicates the most recent calendar year for the publications being considered. The user must edit this variable, but can accept the default values for the other variables.
   (b) TopPub—with a default value of 30, this refers to the number of top cited publications used for evaluating ${}^{30a.adj}R_i$;
   (c) the default value is 15, representing the time interval (in years) before the specified "year" used for evaluating ${}^{15y.adj}R_i$;
   (d) Year2ndInt—this variable has a default value of 5, indicating the time interval (in years) before "year" for evaluating ${}^{5y.adj}R_i$;

(e) NorFun—the default function is set to "sqrt" for normalizing NA. Possible options include "sqrt" (square root), "log" (logarithmic), and "linear" (linear function).
(f) NorFunT—this term, defaulted to "sqrt", standardizes the teaching measure T, with the same options as NorFun.
(g) LastAuthor—with a default value of "always", this indicates whether the last author is included in the NCSFSL count. You can choose "always" to include the last author or "NOalphabet" to count the last name only if it does not follow alphabetical order.
(h) Vector "*r*"—the default values are set to (1, 1, 1, 1, 1, 1, 1) which represent the seven weights ($r_{1,\ldots,7}$) (each r_i must be ≥ 0) to be used in Eq. 1; setting $r = (0, 1, 1, 1, 1, 1, 1)$ converts Eq. 1 to the C-score proposed by Ioannidis et al. (2019, 2020) when the second author is not considered;

(i) vector "*w*"—(default values: 1, 1, 1, 1) indicates the four weights (w, w_2, w_3, w_4) (each w must be ≥ 0) to be used in Eq. 4;
(j) "RT_rw"—With a default value of 0.70, this term represents the fraction $0 \leq RT_{rw} \leq 1$ used in Eq. 6; it determines the weight assigned to the publication score versus the teaching score; for example, 0.70 corresponds to 70% publication and 30% teaching.
(k) "outputfile"—the default output file name is set to "RT_ranking.xlsx".

4. Additionally, the code's first section includes a matrix labeled "File", which the user must edit. Each row represents an individual scholar and must include the following information in the specified order:

(a) "Name" of the scholar;
(b) "Scopus ID" of the scholar;
(c) the name of the "author.csv" downloaded from SCOPUS; write only: Author(s), Year, Citation count;
(d) the 3 scaling factors (α, β and γ) related to the specific academic field, if these coefficients are not necessary, set $\alpha_i = \beta_i = \gamma_i = 1$ for each academics;
(e) "Nh" is the number of teaching hours, setting RT_p = 1.00 excludes this voice from the RT ranking score.

Upon executing the MATLAB code, the "TotScore" sheet will be displayed in the MATLAB Command window, and an Excel file named "RT_method.xlsx" will be generated. This file consists of six sheets, as follows:

- The sheet "TotScore" contains the comprehensive standardized citation scores "R", "T" and "RT" (from 0 to 100) for each person as calculated using Eqs. 4, 5 and 6, respectively, RT is ranked;
- sheet "CurrVitae" contains the standardized citation metrics and scores (0-100) related to the full CVs of each author ${}^{\text{all.adj}}\text{R}_\text{i}$ for the scientists i = 1, ..., N;
- sheet "Top30" contains the standardized citation metrics and scores (0-100) related to ${}^{\text{30a.adj}}\text{R}_\text{i}$ for the scholars i = 1, ..., N;
- sheet "Last15y" contains the standardized citation metrics and scores (0-100) related to ${}^{\text{15y.adj}}\text{R}_\text{i}$ for the scholars i = 1, ..., N;
- sheet "Last5y" contains the standardized citation metrics and scores (0-100) related to ${}^{\text{5y.adj}}\text{R}_\text{i}$ for the scholars i = 1, ..., N;

- sheet "File" contains the matrix named "File" indicated above.

In the four sheets containing standardized citation metrics and scores, the columns are structured as follows:

- Column 1: "Name", names of the scientists;
- Column 2: "N. authors", total number of co-authors in the selected papers for each scientist;
- Columns 3-10: the variables—"NA", "H", "Hm", "NC", "NCS", "NCSF", "NCSFSL", and the relative score "R-Score (0-100)";
- Columns 11-18: the same variables adjusted by the 3 scaling factors (α, β and γ)—"alpha*NA", "beta*H", "beta*Hm", "gamma*NC", "gamma*NCS", "gamma*NCSF", "gamma*NCSFSL", and the relative scaled score "scaled R-Score (0-100)".

Ranking in each matrix is based on the "scaled score" values, which range from 0 to 100.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11192-025-05317-y.

Author Contributions N. Scafetta: Conceptualization, Formal analysis, Writing—original draft, Writing—review & editing.

Funding Open access funding provided by Università degli Studi di Napoli Federico II within the CRUI-CARE Agreement. No funding was received for this work.

Declarations

Conflict of interest The author has no financial interests to disclose. A potential non-financial conflict of interest is acknowledged, as the paper examines criteria for the evaluation of scientific performance within academia, a context in which the author is professionally engaged as a university faculty member. The author affirms that all analyses and conclusions have been conducted with rigorous objectivity to minimize potential bias.

References

Alberts, B. (2013). Impact factor distortions. *Science, 340*(6134), 787.
Bjørk, R. (2019). The age at which Noble Prize research is conducted. *Scientometrics, 119*, 931–939.
Bjørk, R. (2020). The journals in physics that publish Nobel Prize research. *Scientometrics, 122*, 817–823.
Bornmann, L., & Hans-Dieter, D. (2007). What do we know about the h-index? *Journal of the American Society for Information Science and Technology, 58*(9), 1381–1385.
Brembs, B., Button, K., & Munafò, M. (2013). Deep impact: unintended consequences of journal rank. *Frontiers in Human Neuroscience, 7*, 291.

CoARA (2022). Agreement on reforming research assessment. Available at https://coara.eu/, accessed on Feb/07/2025. Download from: https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf

Compagnucci, L., & Spigarelli, F. (2020). The Third Mission of the university: A systematic literature review on potentials and constraints. *Technological Forecasting and Social Change, 161*, 120284.

Gallina, P., & Gallo, O. (2020). Asphyxia of Italian academia in medicine and political deference. *The Lancet, 396*(10247), 307.

Gillett, R. (1991). Pitfalls in assessing research performance by grant income. *Scientometrics, 22*, 253–263.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics, The Leiden Manifesto for research metrics. *Nature, 520*, 429–431.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *PNAS, 102*(46), 16569–72.

Hollanders, H., Tolias, Y., Fabbri, E. et al., European Commission, Joint Research Centre: Measuring territorial innovation strengths, Fabbri, E.(editor), Radovanović, N.(editor), González Evangelista, M.(editor), Publications Office of the European Union, 2024, https://data.europa.eu/doi/10.2760/73871

Horta, H. (2022). Academic inbreeding: Academic oligarchy, effects, and barriers to change. *Minerva, 60*, 593–613.

Ioannidis J.P.A. (2024). Data-update for Updated science-wide author databases of standardized citation indicators. Elsevier Data Repository. Available at: https://doi.org/10.17632/btchxktzyw.7

Ioannidis, J. P. A., Baas, J., Klavans, R., & Boyack, K. W. (2019). A standardized citation metrics author database annotated for scientific field. *PLoS Biology, 17*(8), e3000384.

Ioannidis, J. P. A., Boyack, K. W., & Baas, J. (2020). Updated science-wide author databases of standardized citation indicators. *PLoS Biology, 18*(10), e3000918.

Ioannidis, J. P. A., Klavans, R., & Boyack, K. W. (2016). Multiple citation indicators and their composite across scientific disciplines. *PLoS Biology, 14*(7), e1002501.

Jauch, L. R., & Glueck, W. F. (1975). Evaluation of university professors' research performance. *Management Science, 22*(1), 66–75.

Johnes, J., & Johnes, G. (1995). Research funding and performance in UK university departments of economics: A frontier analysis. *Economics of Education Review, 14*, 301–314.

Laudel, G. (2005). Is external research funding a valid indicator for research performance? *Research Evaluation, 14*(1), 27–34.

Moher, D., Naudet, F., Cristea, I. A., Miedema, F., Ioannidis, J. P. A., & Goodman, S. N. (2018). Assessing scientists for hiring, promotion, and tenure. *PLoS Biology, 16*(3), e2004089.

Nogrady, B. (2023). Hyperauthorship: The publishing challenges for 'big team' science. *Nature, 615*, 175–177.

Noorden, R. V. (2023). More than 10,000 research papers were retracted in 2023 - a new record. *Nature, 624*, 479–481.

Philip, J. (2023). Does C-score-based world ranking give an accurate account of the impact of research papers? *Mendeley Data*. https://doi.org/10.17632/f9gny773br.1

Schotten M., el Aisati M., Meester W., Steiginga S., & Ross C. (2017). A brief history of Scopus: The world's largest abstract and citation database of scientific literature. In Cantu-Ortiz F., Research Analytics. Boosting University Productivity and Competitiveness through Scientometrics.

Schreiber, M. (2008). A modification of the H-index: The Hm-index accounts for multi-authored manuscripts. *Journal of Informetrics, 2*(3), 211–216.

Strathern, M. (1997). Improving ratings: Audit in the British University system. *European Review, 5*(3), 305–321.

Thelwall, M., Kousha, K., Abdoli, M., Stuart, E., Makita, M., Font-Julián, C. I., Wilson, P., & Jonathan, Levitt J. (2023). Is research funding always beneficial? A cross-disciplinary analysis of U.K. research 2014–20. *Quantitative Science Studies, 4*(2), 501–534.

Vieira, E. S., & Gomes, J. A. N. F. (2010). Citations to scientific articles: Its distribution and dependence on the article features. *Journal of Informetrics, 4*(1), 1–13.

Waltman, L., & van Eck, N. J. (2012). The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology, 63*, 406–415.

Wojick, D.E., Michaels P.J. (2015). Is the government buying science or support? A framework analysis of federal funding-induced biases. CATO working paper no. 29. CATO Institute, Washington DC, USA. Available at https://www.cato.org/sites/cato.org/files/pubs/pdf/working-paper-29.pdf, accessed on Feb/07/2025.