Salvatore Chirumbolo. Top Ital. sci. j. 2025;2(1)
https://www.doi.org/10.62684/ANCA8483

Top Italian Scientists Journal

Open Access | Editorial | Scientometrics

ISSN 3033-5132

The meritocratic degree of the Academic evaluation in Italy. Should artificial intelligence be used in this context?

Salvatore Chirumbolo

Department of Engineering for Innovation Medicine, University of Verona, Verona, Italy

Correspondence:

Salvatore Chirumbolo
Department of Engineering for Innovation Medicine
University of Verona, Italy
Strada Le Grazie 8, 37134 Verona
Tel. +390458027645
e-mail salvatore.chirumbolo@univr.it

Keywords: Artificial intelligence, academic evaluation in Italy

Published: 4 Jan 2025

Abstract

The increasing emphasis on publication quantity over quality in academic evaluation has raised concerns about the integrity of meritocratic assessments in Italy. This study explores the impact of Letters to the Editor (LTEs) and Correspondences on reputation indices, such as the h-index, and proposes an algorithm to minimize their disproportionate influence. The proliferation of AI tools, including machine learning and chatbots, has further complicated academic evaluation by enabling rapid manuscript production, potentially inflating publication counts without corresponding research depth. This paper introduces a weighted algorithm that adjusts the h-index by assigning lower weights

to LTEs and Correspondences, based on predefined parameters. Using example datasets, the adjusted h-index significantly reduced scores compared to the original, reflecting more accurate merit assessments. Statistical analysis revealed a high correlation between original and adjusted h-indices (Pearson r = 0.98, Spearman r = 0.94), yet highlighted a weak correlation between letter contributions and rank order (r = -0.04). Paired t-tests confirmed significant differences between the two indices (p = 0.0393). The proposed method effectively penalizes superficial publications while preserving the value of full-length research articles. This approach offers a rigorous and mathematically grounded framework for evaluating academic productivity, particularly in high-stakes settings like Italy's National Scientific Qualification competitions. Furthermore, integrating AI tools into evaluation systems could enhance transparency and fairness, reducing biases and legal disputes related to subjective judgments. Future research should refine these metrics for broader disciplinary applications, ensuring a balanced assessment of academic contributions in the era of artificial intelligence and automated publishing tools.

Editorial

Recently, the Academy where I work publicly acknowledged the prestigious Clarivate achievement of being among the most cited experts in the world for as many as five Full Professors, one of whom has over 22% of publications written in the form of Letters to the Editor or Correspondences—a percentage that accounts for more than one-fifth of the total research output of that author.

One wonders, in truth, whether the reputation of an academic, scholar, or any expert in the scientific field is based more on the number of publications than on their quality. It seems that in Italy, quantity matters as much as quality, especially when reputation carries weight in managerial and political contexts. But how did we get here?

Because, for financial and socio-political impact reasons, universities are increasingly and dangerously resembling corporate structures, with sometimes aggressive financial administrations, where, unsurprisingly, numbers matter, with quality serving as reinforcement. This is one of the reasons why scientific research publications are referred to as "research products"—a term (product) favoured by corporate governance.

As far as is known, under the pretext of participating in the scientific debate within the expert community, it is always possible to open a free and widely accessible online database (for example, the CDC's Data Warehouse WONDER in the United States), conduct research on a specific topic (for instance, the number of deaths from breast cancer during the COVID-19 era), process the results into

a graph, and interpret it to complete a manuscript written in the form of a Letter to the Editor. This then becomes a formal publication, produced very quickly and therefore easy to draft, should one wish to bolster their collection of scientific reports.

While this behaviour could be considered ethical if the expert is genuinely participating in a scientific debate, especially if they are at the beginning of their career or need to embark on an academic path—though many legitimate objections and concerns could be raised—it takes on troubling connotations if such an approach occupies the time of already established Full Professors. This leads one to believe that it might even reflect compulsive tendencies, broadly falling within the category of "publish or perish" or even psychiatric concerns [1,2].

Added to this is the alarming danger stemming from the excessive and deceptive use of machine learning technologies, such as Chat GPT and various artificial intelligence (AI) tools, which could turn this approach into a veritable assembly line of ready-to-use publications, aimed at building a body of work large enough to inflate one's H-index and secure a position in the top global rankings.

However, AI might give some unsuspected positive chance to a serious working in the Academies.

Anyway, to reduce the impact of Letters to the Editors and Correspondences on a reputation index in scientometric evaluations, a mathematically sound algorithm can be developed.

Letters to the Editors (LTEs) and Correspondences are often short publications with limited citation potential, and including them in reputation indices (e.g., h-index, citation count) may disproportionately affect a researcher's evaluation. The goal is to create an algorithm that: a) excludes or de-emphasizes LTEs and Correspondences; b) ensures fairness while preserving the contributions of full-length articles, reviews, and original research papers.

In the proposed algorithm, we can address these parameters and indexes: P = total number of publications; $C_{(i)}$ = number of citations for the i^{th} publication; T(i) = type of the i^{th} publication (e.g., LTE, Correspondence, Review, Research article and so forth); α = weighing factor for LTE and Correspondences ($0 \leq \alpha \leq 1$); β = citation threshold below which LTEs are ignored (optional).

Then, we can assign weights based on publication type: Research articles and Reviews $w_i = 1$; LTEs and Correspondences $w_i = \alpha$ (where $\alpha < 1$). For example, $\alpha = 0.2$ LTEs contribute 20% of the weight of a regular paper.

To adjust the citation count for each paper:

(1) $C^*_{(i)} = w_i \cdot C_{(i)}$

If a citation threshold (β) is used, exclude LTEs with citations below β:

(2) $C^*_{(i)} = 0$ if $T_{(i)}$ = LTE and $C_{(i)} < \beta$

In this perspective, we can estimate the weighed H_{index}, where we could modify the H_{index} using the weighted citation count C*(i). Publications are ranked in descending order of $C^*_{(i)}$ and the weighted index h is:

(3) $h^* = \max \{h: C^*(h) \geq h\}$

To evaluate the Normalized Citation Impact (NCI), we should consider:

(4) $NCI = \frac{\sum_{i=1}^{P} C^*_{(i)}}{P_{eff}}$

where P_{eff} = the effective publication count based on weights:

(5) $P_{eff} = \sum_{i=1}^{P} w_i$

The overall interpretation is that h* reduces the impact of LTEs on the h-index, NCI accounts for the relative contribution of LTEs without inflating metrics. The weights (α) and thresholds (β) can be adjusted for different disciplines or evaluation standards.

If considering exemplificative H index of six randomly selected authors or experts:

Expert	h-index (h)	% Letters (f_l)	Adjusted h-index (h^*)
A	103	22.1% (0.221)	85.64
B	107	13.6% (0.136)	95.68
C	69	5.7% (0.057)	66.04
D	36	1.4% (0.014)	35.29
E	61	5.4% (0.054)	58.88
F	43	33.8% (0.338)	33.19

Consider these assumptions and parameters:

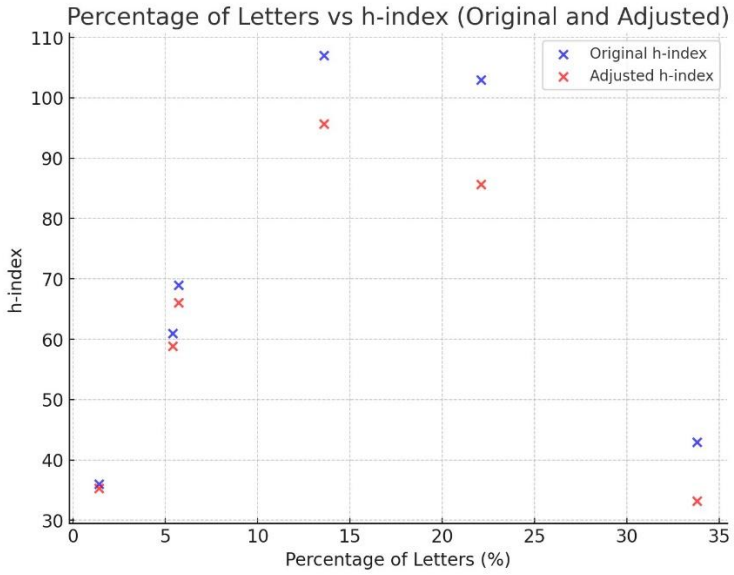
Weight for research articles (w_a) = 1.0

Weight for Letters (w_l) = 0.2

Formula for adjusted H_{index} (h)*: $h^* = h\ (1-f_l + w_l \cdot f_l)$,

where h = original H index, fl = fraction of letters in total publications (as a decimal), w_l = weight for Letters (default 0.2).

Figure 1 shows the adjusted H index following these estimations.



In this plot descriptive statistics reported that the mean original h-index is 69.83, whereas the mean adjusted h-index is 62.45. Furthermore, the percentage of letters varies widely (1.4% to 33.8%). Therefore, adjusted h-index values are lower than original h-index values, indicating a penalty due to letter contributions. In addition, the Pearson's correlation (linear relationships) provides these data: between original h-index and adjusted h-index we have 0.98 (high positive correlation), whereas between percentage of Letters and adjusted h-index we have -0.04 (weak negative correlation). Similarly, the Spearman's correlation (rank-based) resulted that between original h-index and adjusted h-index was 0.94 (strong rank correlation), while between percentage of Letters and adjusted h-index we have 0.09 (weak correlation). The paired t-test gave a t-statistic = 2.77, p-value = 0.0393, assessing that the adjusted h-index is significantly different (lower) than the original h-index at the 5% significance level. The percentage of letters has a weak correlation with h-index values, indicating little influence on rank order. Adjusted h-index values are significantly lower than original values, suggesting a notable effect of accounting for letter weights in the calculation.

This mathematical and quite rigorous evaluation, it should be employed, along with other scientometric variants for the proper evaluation of meritocracy, for example, in evaluation competitions to become a university professor, such as in the National Scientific Qualification competitions, where the use of machine learning, like any chatbot, could even be useful for the accurate and indisputable assessment of the relevance of publications to ministerial declarations, that is, whether the publication is relevant to the profile specified by the Ministry for the specific Scientific-Disciplinary Sector (SSD).

In these cases, the Commission's judgment is very often or almost always highly subjective and possibly even influenced by potential interests or biases, forcing the candidate, unjustly deemed

unsuitable, to defend themselves through legal means, resorting to the costly institution of an appeal. Stricter and less subjective evaluations would ensure the legitimate use and impact of meritocracy.

Declarations

Conflict of Interest

The Author declares that there is no conflict of interest.

Author Contributions

Conceptualization, Data curation, Software, Manuscript writing, reviewing and submitting.

Figure legend

Figure 1. The scatter plot illustrates the trends between the percentage of letters and both original and adjusted h-index values. No strong visual correlation emerges between percentage of letters and h-index, supporting the correlation findings. The percentage of letters has a weak correlation with h-index values, indicating little influence on rank order. Adjusted h-index values are significantly lower than original values, suggesting a notable effect of accounting for letter weights in the calculation.

References

1. Chirumbolo S, Bjørklund G. Pressure to publish in the biomedical scientific field: Ethical conflicts or a possible obsessive-compulsive disorder? Eur J Intern Med. 2018 Apr;50:e16-e17.
2. Chirumbolo S, Bjørklund G. Scientific Pressure to Publish: An OCD or a Gambling Disorder-Like Addictive Behaviour?[Presión científica para publicar: ¿un TOC o un comportamiento adictivo similar al del juego?] Rev Colomb Psiquatria, 2022, doi: 10.1016/j.rcp.2022.10.010