Revista Brasileira de Ciências Ambientais
Brazilian Journal of Environmental Sciences

Bibliometric indicators in the context of the evaluation of science and of Brazilian graduate education: the thermometer measures temperature, not distance

Indicadores bibliométricos no contexto da avaliação da ciência e da pós-graduação brasileira: o termômetro mede a temperatura, não a distância

Valdir Fernandes[1], *Avelino Francisco Zorzo*[2], *Antonio G. Souza Filho*[3]

ABSTRACT

Bibliometric indicators – the JCR, SNIP, SJR, CiteScore, FWCI, and the percentiles and quartiles of the Web of Science and Scopus indexing databases, along with the alternative metrics such as Altmetrics and PlumX – are currently the main instruments used as proxies for assessing the quality and impact of scientific journals and articles. Despite their wide usage, these indicators are frequently misunderstood and often misused – especially when applied to comparisons between fields with structurally distinct sizes and citation culture. This paper discusses the logic underlying their construction, the equations used to calculate them, and the conditions under which these indicators could be used, with emphasis on the specificities of scientific fields with smaller communities – such as the multidisciplinary sciences, the applied social sciences, and the humanities – which may be placed at a disadvantage by metrics calibrated for fields whose articles have high citation density and large collaboration networks. It argues that the established bibliometric indicators are necessary instruments and that, when combined with rigorous indexing processes, they even serve as a defense against journals whose editorial practices do not ensure the integrity of the publication process – referred to in the international literature as predatory. In this context, however, they must be understood within their limits and applied to the field for which they were designed for, within a framework of responsible and contextualized evaluation. In Brazil, this debate takes on the scenario whereby there is a transition from the article classification formerly associated with the CAPES Qualis methodology, centered mainly on journal percentiles, to the three complementary classification procedures established for the evaluation of graduate programs in the 2025-2028 quadrennium. These procedures incorporate both journal-level bibliometric indicators – percentiles, the CiteScore, the JIF, and, alternatively, the H5 – and bibliometric indicators calculated directly at the article level, such as the FWCI. In this new scenario, understanding these instruments ceases to be merely a desirable technical competence and becomes an operational requirement for researchers and graduate program coordinators.

Keywords: Impact Factor; Percentiles; CiteScore; SNIP; Science Evaluation; Graduate Education Evaluation.

RESUMO

Os indicadores bibliométricos — JCR, SNIP, SJR, CiteScore, FWCI, percentis e quartis das bases de indexação Web of Science e Scopus, além das métricas alternativas Altmetrics e PlumX — são atualmente os principais instrumentos utilizados como *proxies* para inferir a qualidade e o impacto de periódicos e artigos científicos. Apesar de sua ampla adoção, esses indicadores são frequentemente mal compreendidos e, sobretudo, mal utilizados — especialmente quando aplicados a comparações entre áreas com dimensões e culturas de citação estruturalmente distintas. Este artigo discute a lógica de construção, as equações de cálculo e as condições em que estes indicadores poderiam ser utilizados, com ênfase nas especificidades de campos científicos de comunidades menores, como as ciências multidisciplinares, as ciências sociais aplicadas e as ciências humanas, que podem ficar em desvantagem por métricas calibradas para áreas cujos artigos têm alta densidade de citações e amplas redes de colaboração. Argumenta que os indicadores bibliométricos consolidados são instrumentos necessários e que, quando aliados a processos de indexação rigorosos, funcionam, inclusive, como uma defesa contra periódicos com práticas editoriais que não asseguram a integridade do processo, denominados, na literatura internacional, predatórios. Nesse contexto, devem, porém, ser compreendidos em seus limites e aplicados no campo para o qual foram projetados, no âmbito de uma avaliação responsável e contextualizada. No Brasil, esse debate ganha contornos específicos com a transição da classificação de artigos antes associada à metodologia Qualis da CAPES, centrada principalmente nos percentis dos periódicos, para os três procedimentos complementares de classificação previstos para a avaliação dos programas de pós-graduação no quadriênio 2025-2028. Esses procedimentos incorporam tanto indicadores bibliométricos do periódico — percentis, CiteScore, JIF e, alternativamente, o H5 — quanto indicadores bibliométricos calculados diretamente no nível do artigo, como o FWCI. Nesse novo cenário, compreender esses instrumentos deixa de ser apenas uma competência técnica desejável e torna-se uma exigência operacional para pesquisadores e coordenadores de programas de pós-graduação.

Palavras-chave: Fator de Impacto; Percentis; CiteScore; SNIP; Avaliação da Ciência; Avaliação da Pós-Graduação.

[1]Universidade Tecnológica Federal do Paraná – Curitiba (PR), Brazil.
[2]Pontifícia Universidade Católica do Rio Grande do Sul – Porto Alegre (RS), Brazil.
[3]Universidade Federal do Ceará – Fortaleza (CE), Brazil.
Corresponding author: Valdir Fernandes – Avenida Sete de Setembro, 3165 – Rebouças – 80230-901 – Curitiba (PR), Brazil. E-mail: valdir.fernandes@icloud.com
Received on: 06/30/2026. Accepted on: 08/22/2026.
https://doi.org/10.67790/rbciamb.026.3188

Introduction

Thousands of journals worldwide receive new bibliometric indicators every year. In June 2026, for example, Scopus[1] updated the CiteScore of all journals indexed in its database. Around the same time, Clarivate[2] updated the Journal Citation Reports (JCR) with 2025 data. On the basis of this information, editors update their websites, authors revise their publication strategies, and some institutional evaluation committees consult spreadsheets of impact factors. And a considerable share of these actors does so without fully understanding what the numbers mean, how they were calculated, and — most importantly — the limits of their application and when they should not be compared with one another.

We argue that this comprehension gap is not trivial. Bibliometric indicators are today simultaneously the most widely used and the most poorly understood instruments in research evaluation, and this contradiction has real consequences for researchers, journals, and policy makers for the evaluation of science and of graduate education. Although, as Oliveira Jr and Zorzo (2026) argue, the evaluation of science and of graduate education should not be restricted to bibliometric indicators, when used correctly, they constitute objective and auditable references, useful for distinguishing established journals from opportunistic ones, for assessing the relative performance of journals within their respective fields, and for guiding editorial and funding decisions. Some of these bibliometric indicators, such as the Impact Factor (JIF) and the H5, when used incorrectly — especially when applied to comparisons between fields with radically distinct citation culture — produce distortions that systematically and unfairly penalize entire fields. The great challenge is to use metrics responsibly and to explore the meaning of bibliometric indicators in different evaluation contexts. Bibliometric indicators should not be ending points, but rather starting points for a responsible classification of journals and articles. It should be made clear that the goal of this article is not to determine what indicator is "better", but to discuss the context in which each one can be used — and when its use is inappropriate.

The problem, it is worth stressing from the start, does not lie with the indicators themselves. The Impact Factor does not err in measuring what it sets out to measure; the error lies with those who use it to measure what it was not designed to measure. The thermometer is not at fault when someone uses it to measure geographic distances. This distinction — between the legitimate limitation of an instrument and its inappropriate use — is the guiding thread of this paper.

Scientific journals remain the main means by which peer-validated knowledge is recorded, communicated, and preserved. The accelerated growth in the number of journals in recent decades — driven by digitization, the open-access model, and institutional pressure to publish — has brought with it a worrying phenomenon: the proliferation of journals with questionable editorial practices, known as predatory journals, which simulate the processes of legitimate science without actually carrying them out (Grudniewicz et al., 2019). In this context, understanding the indicators that distinguish legitimate journals from so-called predatory ones is not a secondary technical matter: it is an emerging issue of scientific integrity, one that requires research institutions and researchers to adopt good practices in the dissemination of results and in the training of students.

This article presents the construction logic and the appropriate conditions of usage of the main bibliometric indicators available[3] — the JIF, the SNIP, the SJR, the CiteScore, the FWCI, the percentiles and quartiles of Scopus and the Web of Science (WoS), and the alternative metrics Altmetrics and PlumX — with particular attention to the specificities of the multidisciplinary sciences (such as environmental sciences or teaching), the applied social sciences (such as management or law), and the humanities (such as education or philosophy), although the discussion may also extend to linguistics, literature, and the arts, or to the exact sciences, such as computer science or the geosciences.

Methodologically, this article is a narrative review (a theoretical-analytical essay) of bibliometric indicators, not an empirical study. The selection of the literature prioritized four axes: (i) the original technical articles that define each indicator (e.g., González-Pereira et al., 2010, for the SJR; Moed, 2010, for the SNIP); (ii) the official documentation of the databases that calculate them (the Elsevier Research Metrics Guidebook; Clarivate reports on the JCR); (iii) the milestones of the responsible research assessment movement — the San Francisco Declaration (DORA, 2013), the Leiden Manifesto (Hicks et al., 2015), The Metric Tide report (Wilsdon et al., 2015), and CoARA (2022); and (iv) the official Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES) documentation on the new classification system for assessing the publication in journals for the 2025–2028 quadrennium.

Main indicators

Before presenting the indicators, it is worth distinguishing four concepts frequently treated as synonyms in this debate. "Quality" refers to the intrinsic scientific value of a work or journal, assessed primarily through peer review — something that bibliometric indicators do not measure directly, but only infer by means of proxies. "Impact" refers to the number of citations received, a proxy for use and influence, serving as an indicator of dissemination within the scientific community, but not necessarily of quality. "Prestige" refers to the weight attributed to a citation as a function of the reputation of the citing source — this is what indicators such as the SJR seek to capture, unlike raw impact indicators such as the JIF or the CiteScore.

[1] https://www.scopus.com/sources

[2] https://www.webofscience.com/

[3] Like the FWCI, the InCites Journal Normalized Citation Impact measures article-level impact, but benchmarks it against the average citation rate of the journal itself, in the same year and for the same document type, rather than against the field.

"Attention", finally, refers to the reach and visibility of a work beyond formal academic citation — mentions on social media, in the press, and in public policy documents — captured by alternative metrics such as Altmetrics and PlumX. Throughout this article, each indicator is situated in relation to the concept it is constructed to measure.

Journal Citation Reports (JCR)

The JCR is produced by Clarivate Analytics from the data of the WoS, one of the most rigorous and selective indexing databases in the world. To be indexed in the WoS, a journal undergoes a careful editorial evaluation process that includes regularity of publication, the quality of the peer-review process, the diversity of the editorial board, and compliance with ethical standards.

The main indicator in the JCR is the JIF, which represents the average number of citations received by a journal's articles in the two years preceding the measurement. Conceived by Eugene Garfield in 1955 and formalized in 1972 (Garfield, 1972, 2006), the JIF was originally proposed as a tool to assist university libraries in selecting journal subscriptions. Its logic is transparent and replicable, which contributed to its wide adoption as a proxy for gauging editorial quality.

How is the Impact Factor calculated?

The JIF is calculated using Equation 1:

$$IF(Y) = \frac{C(Y)}{A(Y-1)+A(Y-2)} \quad (1)$$

where:

C(Y) is the total number of citations received in year Y by the articles published in years Y−1 and Y−2

A(Y−1) and A(Y−2) are the numbers of articles published in the two years preceding the year for which the impact factor is calculated.

Example: A journal published 50 articles in 2023 and 65 in 2024. In 2025, these 115 articles were cited 400 times. 2025 JIF = 400 ÷ 115 = 3.48.

In calculating the JIF, the denominator comprises documents that Clarivate classifies as citable items, primarily original and review articles. Editorials, letters, and other non-citable documents are normally not included in the denominator, although any citations such content receives do contribute to the indicator's numerator — an asymmetry that can artificially inflate the JIF of journals that publish a large volume of this type of content, as is often the case for high-circulation journals. Figure 1 illustrates the construction of the indicator.

The JIF, like any bibliometric indicator, has limitations that must be weighed. As Garfield (2006) pointed out, the JIF was conceived as an aggregate measure of journal citations and not as an indicator of the quality or impact of individual articles. In this regard, Larivière and Sugimoto (2019) observe that the distribution of citations among a journal's articles is typically skewed, so that the indicator's mean value may not adequately represent the performance of all published documents. This skewness was quantified by Larivière et al. (2016): in a study of 11 journals, about 70% of the articles did not reach a number of citations equivalent to the journal's JIF within the two-year window used in the calculation. In a journal with a JIF of 4.70, for example, most articles will not reach 5 citations in that period. Consequently, one should not infer that the journal's average performance

Figure 1 – How the JIF is calculated

necessarily reflects each article's impact — a statistically unwarranted conclusion. On the other hand, as Waltman and Traag (2021) argue, this limitation does not invalidate the indicator; rather, it reinforces the need to understand its premises, restrictions, and methodological nuances. Moreover, the indicator preserves its auditability: the articles and citations that enter its calculation can be identified and analyzed individually, thus enabling complementary assessments that reveal the underlying citation distribution and the actual contribution of the most influential articles to the indicator's value. For these reasons, the proper interpretation of the Impact Factor is a key premise for responsible research evaluation.

The correct use of the Impact Factor

The JIF is a robust and well-grounded indicator for measuring a journal's impact within its field and relative to journals with a similar publication profile (subject matter and community). Its use becomes problematic when applied outside that context — especially when the JIFs of journals from fields with very distinct citation culture are compared directly (Seglen, 1997).

Fields such as oncology, artificial intelligence, biochemistry, and particle physics have large scientific communities, very high publication rates, and extensive reference lists — which naturally lead to higher JIF values. Articles in mathematics have few references, which typically leads, when compared with the fields mentioned above, to low impact factor values in the best mathematics journals. Multidisciplinary fields, such as the environmental sciences, the geosciences, and teaching, operate with smaller communities and shorter reference lists. An excellent journal in these areas may have a JIF lower than that of an average biomedical journal — not because it is of lower quality, but because the citation dynamics of the two fields are incomparable (Seglen, 1997; Waltman, 2016). In short, the JIF should not be ignored and can be important for all fields, provided it is used for within-field comparisons and as a starting point for deeper evaluations, complemented by other indicators and analyses.

Source Normalized Impact per Paper (SNIP)

The SNIP[4] was developed by Henk Moed at the Centre for Science and Technology Studies (CWTS) of Leiden University and implemented in the Scopus database (Colledge et al., 2010; Moed, 2010). Its central methodological proposal is to introduce a correction based on the citation potential of each field of knowledge, thereby enabling comparisons between journals from fields with distinct citation cultures. In the WoS, the conceptually closest indicator is the Journal Citation Indicator (JCI), launched by Clarivate in 2021 as a field-normalized measure of citation impact, complementary to the JIF (Clarivate, 2021a).

How is the SNIP calculated?

The SNIP is calculated using Equation 2:

$$SNIP = \frac{RIP}{RDCP} \quad (2)$$

where:

RIP = Raw Impact per Paper (average citations per paper, 3-year window)

RDCP = Relative Database Citation Potential (the citation potential of the field, also called Citation Potential in the Elsevier guidebook; Moed, 2010) — not to be confused with the CWTS's MNCS (Mean Normalized Citation Score), a distinct, article-level indicator

How does the SNIP equalize impact across fields?

To understand the SNIP's normalizing effect, consider two hypothetical journals: one in molecular oncology and another in the environmental sciences. Oncology articles are typically cited between 30 and 50 times in the three years following publication; environmental science articles rarely exceed 10 to 15 citations over the same period. If the citation potential of molecular oncology is 42 and that of the environmental sciences is 11:

Journal A (oncology): SNIP = 38 ÷ 42 = 0.90

Journal B (environmental sciences): SNIP = 10 ÷ 11 = 0.91

Despite the differences in citation counts, the SNIP values are practically identical — both journals occupy equivalent positions in terms of relevance within their respective fields.

By considering impact into perspective against the citation behavior characteristic of each field, the SNIP makes it possible to state, on methodological grounds, that two journals with similar SNIP values may be equally representative in their fields — even if their JIFs are radically different. It is worth noting that the SNIP currently published in Scopus corresponds to the revised version of the indicator (Waltman et al., 2013), rather than the original formulation proposed by Moed (2010). Figure 2 illustrates the SNIP calculation process.

Scimago Journal Rank (SJR)

The SJR[5] is calculated by the Scimago Group at the University of Granada using Scopus data. Its methodology was formally described by González-Pereira et al. (2010), refined by Guerrero-Bote and Moya-Anegón (2012), and represents one of the most interesting contributions of recent bibliometrics: the incorporation of the concept of the prestige of citing sources — shifting the question from "how many times is a journal cited?" to "by which journal is it cited?"

The SJR is based on the PageRank algorithm — the same one used by Google to rank web pages (Page et al., 1999; González-Pereira et al., 2010). Being cited by a widely recognized journal that is well embedded in research networks confers greater value than being cited by a journal with limited circulation. Larivière and Sugimoto (2019)

[4] https://www.journalindicators.com

[5] https://www.scimagojr.com/

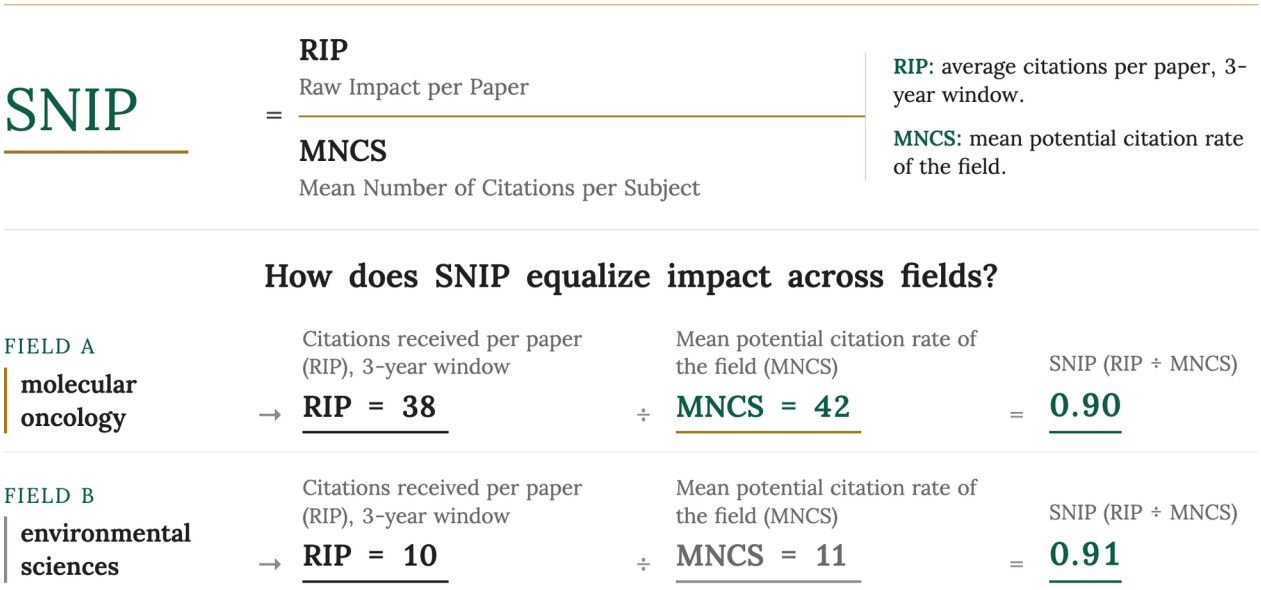

Figure 2 – How the SNIP is calculated

point out that the SJR applies explicit safeguards against manipulation — such as limits on self-citation and on the concentration of prestige among a few journals — which makes it more robust to distortions than indicators based on raw citation counts only. In WoS, the conceptually equivalent indicators are the Eigenfactor Score and the Article Influence Score, both of which are based on a prestige-weighted citation network (Bergstrom, 2007).

How is the SJR calculated?

The SJR of a given journal i is calculated in two phases, in the SJR2 version — the 2012 revision that updated the indicator's original formulation (Guerrero-Bote and Moya-Anegón, 2012) and is the version in use ever since in Scimago and Scopus, where the label remains "SJR". The first phase, expressed in Equation 3, iteratively computes the journal's raw prestige (PSJR2):

$$PSJR2_i = \frac{1-d-e}{N} + e\frac{Art_i}{\sum_j Art_j} + \frac{d}{PSJR2_D}\sum_j Coef_{ji} PSJR2_j \quad (3)$$

The second phase converts the raw prestige into the final indicator, independent of journal size (Equation 4):

$$SJR2_i = \frac{\frac{PSJR2_i}{Art_i}}{\frac{\sum_j Art_j}{}} \quad (4)$$

where:

d = a constant (equal to 0.9)

e = a constant (equal to 0.0999)

N = number of journals in the database

Art_i = number of citable documents (articles, reviews, *short surveys*, and conference papers) published in journal i in the previous 3 years; $\Sigma_j Art_j$ = the total of such documents across all journals in the database

C_{ji} = citations from journal j, in the year of analysis, to documents of journal i published within the 3-year window

Cos_{ji} = cosine between the cocitation profiles of journals j and i — the measure of thematic closeness between them

$Coef_{ji} = (Cos_{ji} \cdot C_{ji}) / \Sigma_h (Cos_{jh} \cdot C_{jh})$ = fraction of journal j's prestige transferred to journal i, calculated before the iterations begin; the transfer — to other journals or to itself — is capped at 50% of the source journal's prestige and at 10% per citation

$PSJR2_j$ = prestige of the citing journal j, obtained in the previous iteration

$PSJR2_D$ = total prestige distributed through citations in the iteration, recalculated at each step; as the divisor, it redistributes the prestige of journals with no outgoing citations (*dangling nodes*) in proportion to the prestige each journal receives

Since the raw prestige is divided by the journal's proportion of citable documents (Equation 4), the weighted average of SJR values is, by construction, equal to 1: values above 1 indicate above-average prestige per document, and values below 1 the opposite — the same interpretive logic as the SNIP and the FWCI.

Figure 3 conceptually illustrates the dynamics of prestige transfer between journals in the citation network used for calculating the SJR indicator. The constants d and e are not readjusted at each calculation: they are fixed parameters, defined by the authors of the algorithm (González-Pereira et al., 2010), and maintained in the 2012 revision, with d = 0.9 and e = 0.0999. The constant d weights the

prestige component transferred through citations — functioning as the damping factor of PageRank, whose analogous value is 0.85 (Brin and Page, 1998) — whereas e weights the component associated with the number of articles published by the journal. The remaining term (1 – d – e) = 0.0001 corresponds to the minimum prestige assigned to any journal simply for being indexed in the database, distributed equally among the N journals, that is, (1 – d – e)/N. These values were calibrated by the authors over the universe of Scopus journals so that the citation-based prestige component predominates, thereby preserving the stability and convergence of the iterative process.

CiteScore

The CiteScore is the central journal evaluation indicator of the Scopus database, developed by Elsevier and published annually. It differs from the Impact Factor in two methodologically relevant ingredients: the time window of the calculation and the document types considered.

How is the CiteScore calculated?

The CiteScore for year Y is calculated using Equation 5:

$$\text{CiteScore}(Y) = \frac{\text{number of citations received in Y–3 to Y by documents published in Y–3 to Y}}{\text{number of documents published between Y–3 and Y}} \quad (5)$$

The time window considered in the CiteScore calculation is four years, including the article's year of publication, as opposed to two years for the JIF. Since the methodological revision of June 2020, both the numerator and the denominator consider exclusively five peer-reviewed document types — research articles, reviews, conference papers, book chapters, and data papers — excluding letters and editorials. This eliminates the asymmetry between numerator and denominator present in the JIF, which counts citations to any document type in the numerator but restricts the denominator to citable items.

The four-year window makes the CiteScore more suitable for fields in which the citation maturation cycle is slower — as is the case of the multidisciplinary sciences, in which long-term studies frequently accumulate citations over many years, a phenomenon documented by Moed et al. (1998).

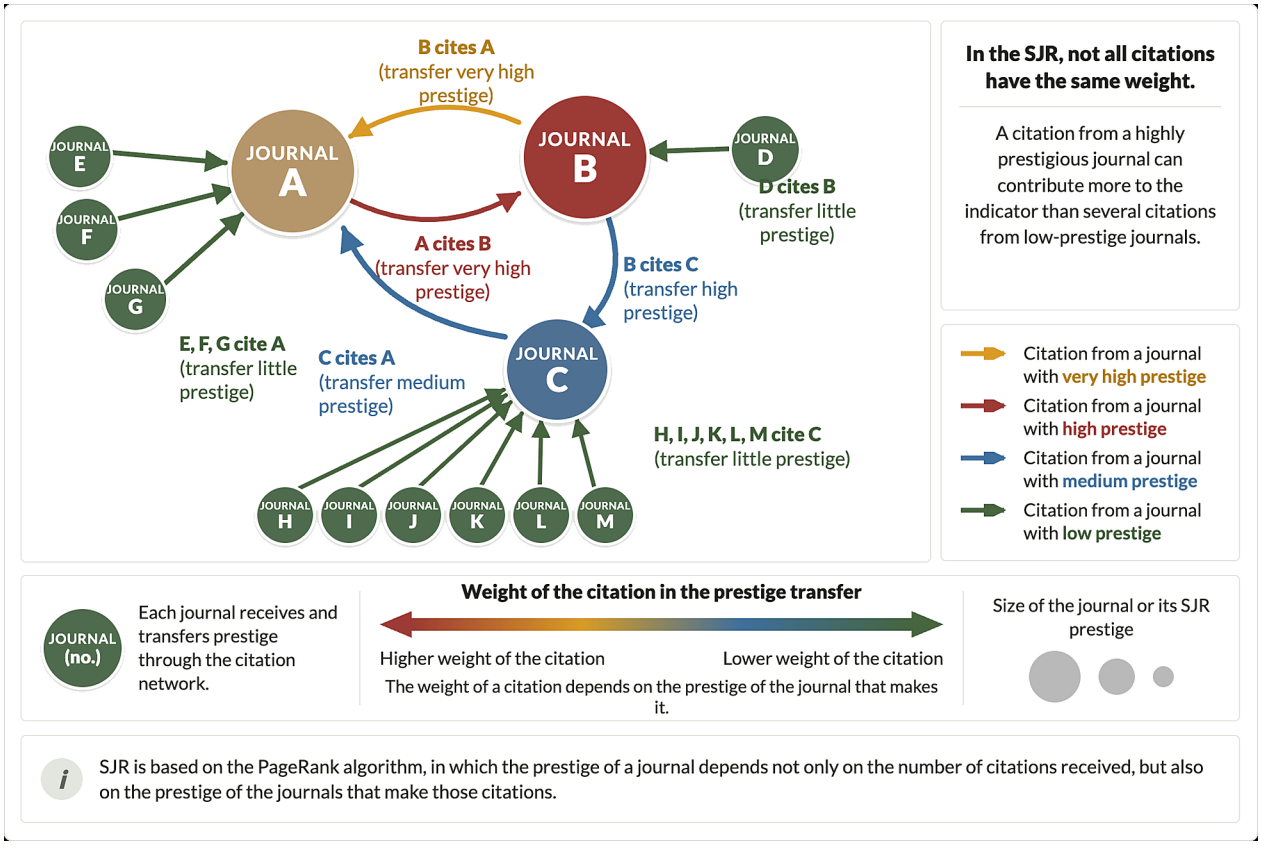

Figure 3 – Conceptual model of PageRank prestige transfer
Source: Prepared by the authors, based on the principles of the PageRank algorithm proposed by Page et al. (1999), adapted to the context of prestige transfer between scientific journals.

The CiteScore is available through the Scopus Sources portal, which also calculates, for each journal, the percentile in each subject category defined by the database itself. It should not be confused with the quartiles published by the Scimago Journal & Country Rank, which derive from the SJR — a distinct ranking system that frequently diverges from the CiteScore of the same journal. Figure 4 illustrates the construction of this indicator, which is the ratio between the citations received and the documents published in the same period.

Field-Weighted Citation Impact (FWCI)

The Field-Weighted Citation Impact (FWCI) is a bibliometric indicator available on the Scopus platform and in Elsevier's SciVal system. Unlike the Impact Factor, the SNIP, and the CiteScore — which measure journal performance — the FWCI is calculated primarily at the article level and can be aggregated at the journal, institutional, national, or publication-set levels (Elsevier, 2019). Its core logic is to compare the citations actually received by a publication with the expected number of citations for documents of the same type, published in the same year, and classified in the same subject area (Waltman and van Eck, 2013). In the WoS, the directly equivalent indicator, calculated by Clarivate's InCites platform, is the Category Normalized Citation Impact (CNCI), which uses the same observed/expected ratio logic and the same aggregation criteria (Clarivate, 2024).

In Dimensions, the Digital Science database, the equivalent indicator is the Field Citation Ratio (FCR): it divides the citations received by a document by the average citations of documents published in the same year and classified in the same Field of Research (FoR) — categories assigned by machine learning, as discussed in Section "Scientific fields with smaller communities: where the inappropriate use of indicators is most costly". Two technical differences deserve mention. First, unlike the FWCI and the CNCI, which aggregate using the arithmetic mean of the individual ratios, the FCR uses the geometric mean of the citation values (Thelwall and Fairclough, 2015), which reduces the influence of highly cited outlier articles. Second, the calculation of the FCR requires a minimum of 500 articles classified in the FoR category for the year in question; below this threshold, the indicator is simply not calculated — a restriction that may leave precisely the smaller-community fields discussed in this article without an FCR.

How is the FWCI calculated?

The FWCI is calculated using Equation 6:

$$FWCI = \frac{CR}{CE} \tag{6}$$

where CR represents the number of citations received in the year and CE represents the expected citations, estimated on the basis of the world average for documents of the same type, publication year, and subject field.

In practice, the CE of a document is obtained by averaging the citations received, within the same four-year window, by all other Scopus-indexed documents published in the same year, of the same document type, and classified in the same ASJC category; when the document is classified in more than one category, the harmonic mean of the averages calculated for each category is used.

FWCI = 1.00 → equal to the world average

FWCI > 1.00 → above average

FWCI < 1.00 → below average

When the FWCI is aggregated to an entity — a journal, institution, country, or research group — the value does not correspond to

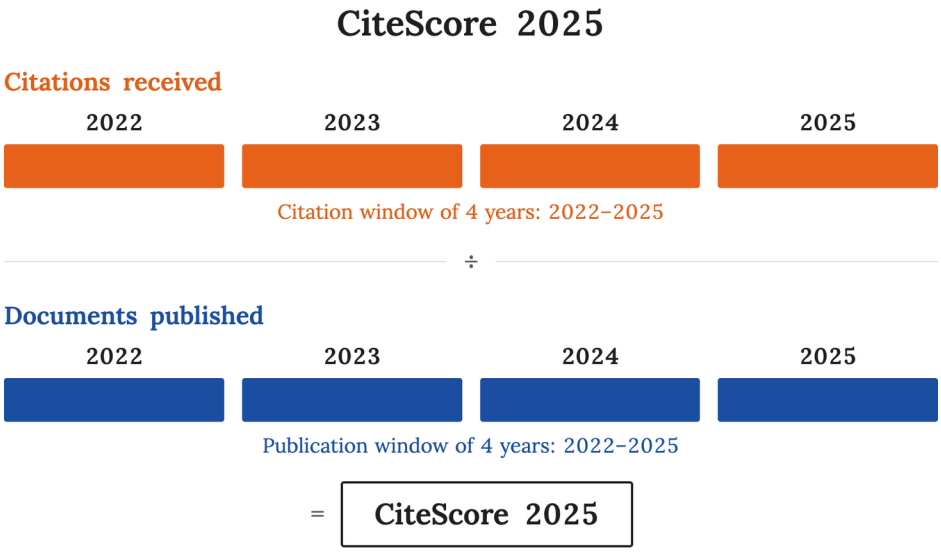

Figure 4 – How the CiteScore 2025 (Scopus) is calculated

the ratio between the total citations received and the total expected citations of the set, but rather to the arithmetic mean of the observed/expected ratios of each individual publication (with the harmonic mean of the expected values when the document is classified in more than one subject category) — a formula confirmed by Elsevier's own guidebook (Elsevier, 2019): FWCI ≡ (1/N)·Σci/ei. In this expression, ci is the number of citations received by publication i and ei is the number of citations expected for that publication, that is, the average citations of documents of the same type, publication year, and subject category; N is the number of publications of the entity. It is worth noting that Elsevier's simplified public documentation (the Scopus Support Center FAQ) at times describes the aggregated FWCI more informally, as the "ratio between the total citations received and the total expected," which would suggest a ratio of sums; the technical definition in the guidebook, however, confirms that it is the mean of the individual ratios, not the ratio of the totals.

An article with an FWCI of 1.50 was cited at 50% above the level expected for equivalent publications worldwide. By normalizing simultaneously for document type, publication year, and subject field, the FWCI controls for three of the main sources of bias affecting comparisons of absolute citation counts: the advantage of older documents, the advantage of fields with higher publication density, and the difference between original articles and systematic reviews. For this reason, the FWCI allows better comparisons of impact across documents, researchers, institutions, and countries, regardless of differences in field, year, or publication type. Figure 5 illustrates how the FWCI compares a document's citations with the expected world average for documents in the same subject field, publication year, and document type.

Applications

The FWCI can be calculated at different levels of aggregation. At the article level, it identifies high-impact publications within a portfolio. At the journal level, it offers a normalized alternative to the Impact Factor for comparing journals across distinct fields — expressed in intuitive terms (as a ratio relative to the world average). At the institutional and national levels, it is widely used in university rankings and in science policy reports, such as the Leiden Ranking[6].

For fields such as the multidisciplinary sciences, the applied social sciences, and others with smaller communities, the FWCI is particularly useful: by measuring performance relative to the field's own average, it avoids comparisons with high-citation-density areas. A field with an FWCI of 1.20, for example, performs 20% above the world average.

Figure 6 illustrates the application of the FWCI, calculated in 2026, at different levels of aggregation. While Brazil has an average FWCI of 0.87, Institutions A and B reach values close to 1.05, and Institution C reaches 1.24, which indicates a scientific impact 24% higher than expected for equivalent publications. Institution D, in turn, has an FWCI of 0.76. These results demonstrate the usefulness of the indicator for comparing countries, institutions, and research groups against the

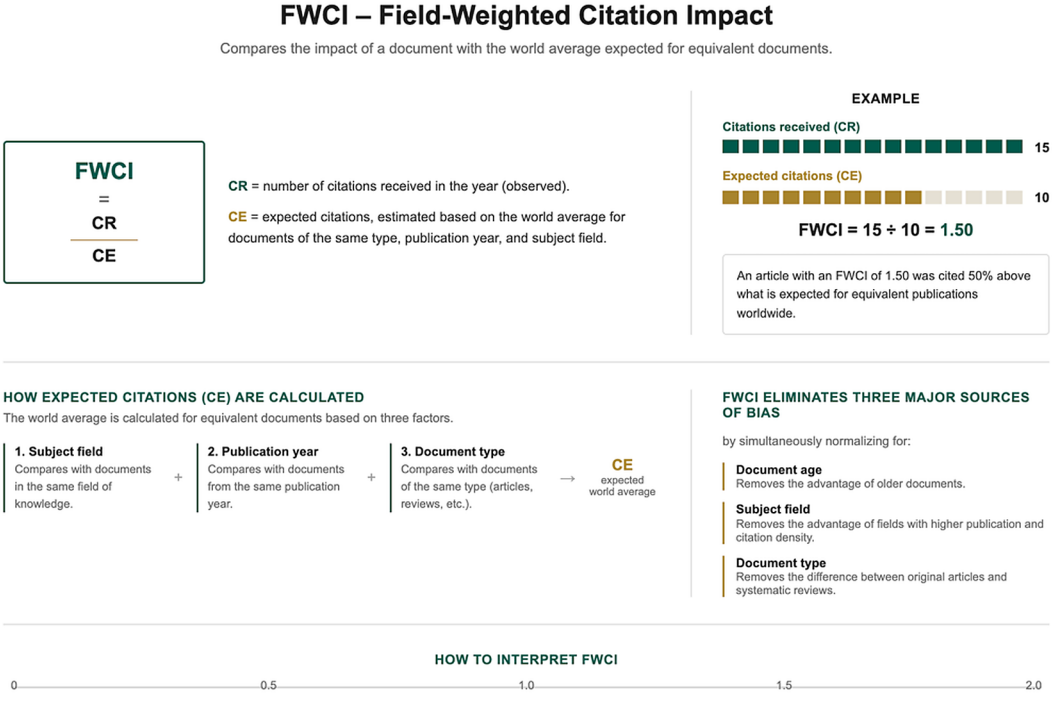

Figure 5 – FWCI infographic

[6] https://www.leidenranking.com/

world average, regardless of differences among fields of knowledge, document types, and publication years (SciVal, 2026).

A complementary feature, not yet mentioned, is fractional counting — especially relevant when the FWCI is aggregated at the level of institutions, countries, programs, or research groups. Field normalization already corrects for differences in citation density across areas, but it does not correct for differences in collaboration practices: fields with typically multi-author, multi-institution, or multinational output can have their aggregate impact inflated when each participant receives full credit for the same article. Dividing this credit among the units involved — the logic of fractional counting — makes comparisons between fields with very different co-authorship profiles more robust. This feature does not replace normalized indicators, such as percentiles, quartiles, or the FWCI itself; it complements them as an additional layer of correction, useful precisely when the unit of analysis is no longer the article or the journal but an aggregated entity.

Percentiles and Quartiles — JCR and Scopus

While the Impact Factor, the SNIP, the SJR, and the CiteScore indicate the performance of a journal through absolute or normalized values, and the FWCI measures impact at different levels of aggregation, such as articles, groups, institutions, or countries, percentiles and quartiles adopt a relative perspective: they indicate the journal's position relative to all others in the same subject field. They are among the most informative instruments for the comparative evaluation of journals precisely because they translate performance into position — and not merely into the magnitude of the indicators. It is important to note that subject groupings vary somewhat from platform to platform, as they depend on how each database indexes journals across subjects.

Both WoS (via the JCR) and Scopus (based on CiteScore) provide percentiles for the journals they index. A journal in the 90th percentile is among the 10% most-cited in its subject category; in the 75th percentile, it outperforms 75% of journals in its field.

Percentiles and quartiles in the Web of Science (JCR)

In the JCR, percentiles are calculated within each subject category based on the Impact Factor. Journals are grouped into quartiles: Q1 (percentiles 75–100), Q2 (50–75), Q3 (25–50), and Q4 (0–25) — widely used in institutional evaluations, competitive examinations for academic positions, and scientific CVs in Brazil and abroad.

The use of JCR quartiles and percentiles largely resolves the problem of comparison across distinct fields: instead of contrasting the absolute JIF value of an environmental science journal with that of a physics journal, one compares the relative position of each in its respective field. A journal in Q1 of environmental sciences is as well positioned in its field as a journal in Q1 of computer science is in its own.

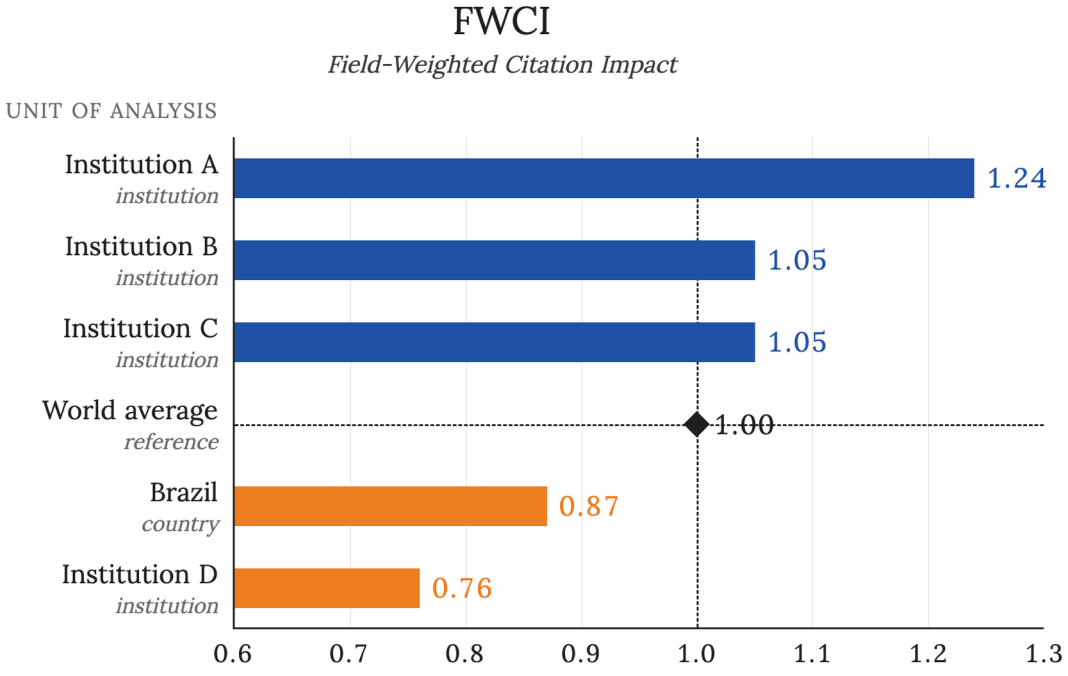

Figure 6 – Examples of the application of the FWCI at different levels of aggregation
Note: Anonymized institutions (Institutions A-D). The dashed line marks the world average (FWCI = 1.00); values above it indicate citation impact higher than expected for the field.
Source: SciVal (2026).

Percentiles and quartiles in Scopus (CiteScore)

In Scopus, the CiteScore percentiles for each subject category are published by Scopus on the Scopus Sources portal. As mentioned earlier, these should not be confused with the quartiles freely available at scimagojr.com, which derive from the SJR — a distinct ranking system calculated by the Scimago Group that may position the same journal differently from CiteScore. The same journal may appear in different subject categories and occupy distinct positions in each — which is especially relevant for interdisciplinary journals, such as those in the environmental sciences, evaluated simultaneously in categories such as Environmental Science, Water Science and Technology, and Ecology. This assignment to multiple categories is not automatic: in general, it results from the journal's request to the indexing database for the subject areas of interest at the time of registration, and it may be revised later at the journal's request. Percentiles indicate the relative position of a journal within its subject category, whereas quartiles group journals into four equal performance bands. Figure 7 presents the percentile scale and the positioning of the quartiles along it.

Altmetrics and PlumX — attention and engagement metrics

All the indicators discussed so far are based on citations among indexed scientific journals. In recent decades, the debate on research evaluation has come to incorporate an additional dimension: impact beyond the academic world, measured through so-called alternative metrics (altmetrics). The concept was formalized by Priem et al. (2010) in the Altmetrics Manifesto, which argues that citations in journals capture only a fraction of the real impact of scientific research (Bornmann, 2014). These metrics seek to capture a broader social dynamic, in which different segments of society use new media and social networks to disseminate, read, and debate a range of topics, including science.

Altmetrics

The Altmetrics system[7] assigns each article an Altmetric Attention Score (AAS), calculated by an algorithm that sums fixed scores assigned to each type of mention — for example, a news story in the press or a public policy document is worth more points than a share on a social network. The result is a raw, cumulative score, not normalized by field of knowledge or by the user volume of the source platform, which limits direct comparisons between fields with different levels of media visibility. For journals in the multidisciplinary sciences, the AAS is particularly informative in research on water quality, deforestation, or extreme climate events, which are frequently cited in documents of bodies such as the UN, the IPCC, and the WHO.

PlumX Metrics

PlumX, integrated into the Scopus platform, organizes the data into five categories:

- **Citations:** references to indexed academic publications.
- **Usage:** views, downloads, and clicks.
- **Captures:** additions to reading lists and exports in reference managers such as Mendeley and Zotero.
- **Mentions:** references in blogs, news outlets, and Wikipedia.
- **Social media:** shares on X (Twitter), Facebook, and LinkedIn.

Figure 8 presents the Altmetric architecture. While the Altmetric Attention Score focuses mainly on sources of online attention, PlumX Metrics also incorporates dimensions of academic impact, engagement, and social influence. This integration of academic and non-academic citations is particularly promising for comparative purposes. In PlumX, the indicators are officially grouped into five categories: Citations, Usage, Captures, Mentions, and Social Media. In Figure 8, these categories have been aggregated into broader conceptual dimensions to facilitate the understanding of the different types of impact and attention generated by scientific output.

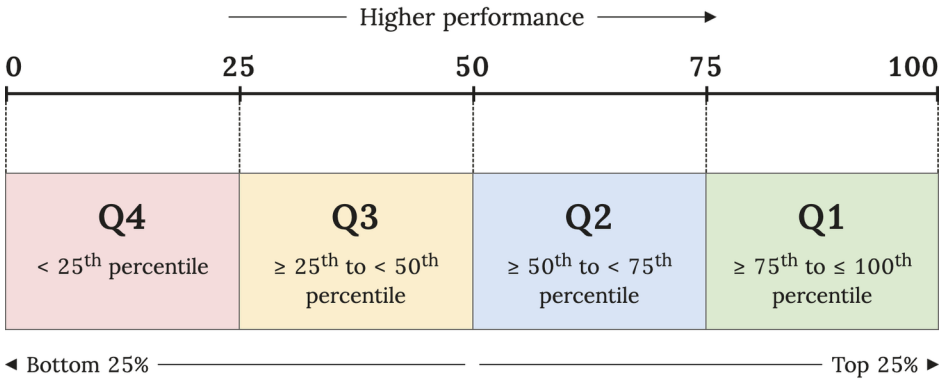

Figure 7 – Percentile scale and the positioning of the quartiles along it

[7] altmetric.com

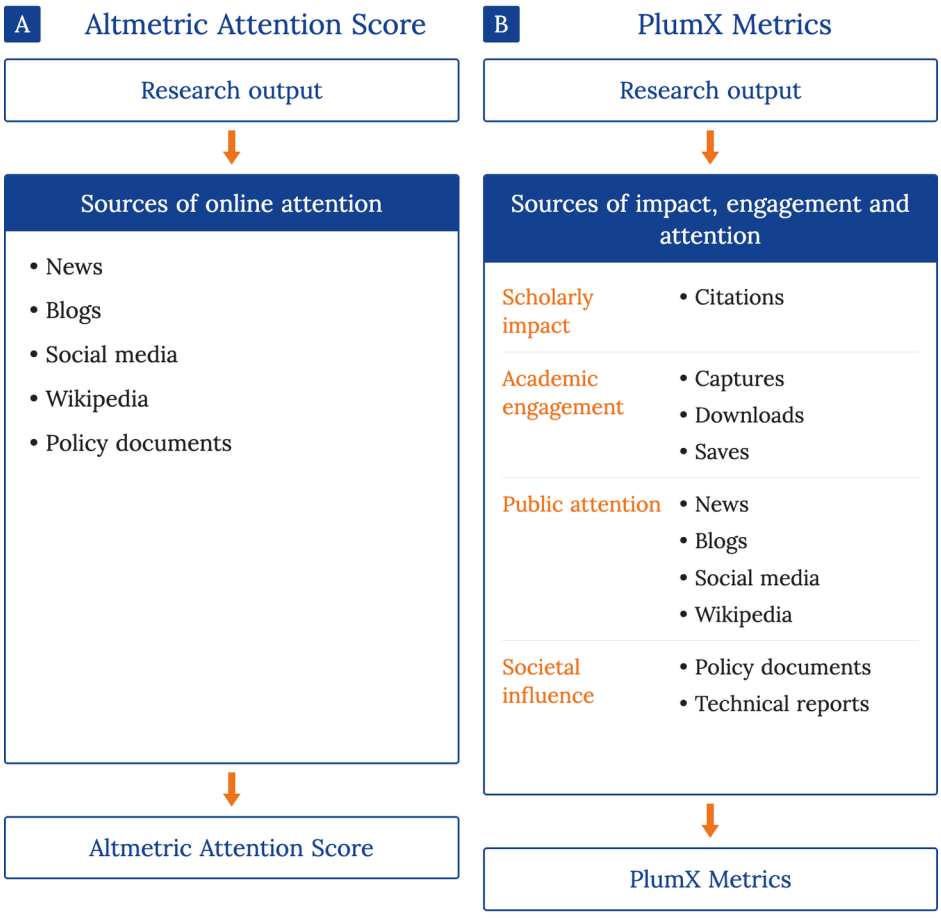

Figure 8 – Main sources of attention, engagement, and impact captured by the Altmetric Attention Score (A) and by PlumX Metrics (B)

Limitations and complementarity

Both Altmetrics and PlumX should be interpreted as complementary instruments — not substitutes — for the indicators already understood and used by the scientific community. Thelwall et al. (2013) demonstrated that the correlation between online attention and academic citations is positive but weak. Bornmann (2014) notes that attention on social media is volatile and susceptible to factors unrelated to the scientific quality of the article (Eysenbach, 2011). Moreover, it is known that the dynamics of engagement on social media depend on algorithms, which are managed with broad autonomy by big tech companies. Table 1 summarizes the main characteristics — database, time window, normalization, and weighting — of all the indicators discussed so far, for quick comparative reference.

Scientific fields with smaller communities: where the inappropriate use of indicators is most costly

Bibliometric indicators were conceived, calibrated, and extensively validated in high-volume scientific fields: biomedicine, chemistry, physics, and engineering. The problem arises when these same indicators are applied — without adjustment or contextualization — to fields that operate under structurally different conditions: smaller communities, slower citation rhythms, multilingual output, and a strong orientation toward local or regional public policy.

These fields include the multidisciplinary sciences, the applied social sciences, the humanities, regional studies, and much of applied engineering. Leydesdorff and Rafols (2011), analyzing the 8,207 journals of the JCR, showed that interdisciplinary journals are distinguished by citation networks that are more diverse and more dispersed across domains than those of specialized journals; Rafols et al. (2012), in an empirical comparison between innovation studies and business and management, showed that this structural diversity translates into a concrete disadvantage: excellence-based journal rankings exhibit a systematic bias in favor of disciplinary research, penalizing the evaluation of interdisciplinary units and researchers.

Multidisciplinarity and citation dispersion

Fields with smaller communities also tend to be more multidisciplinary. An environmental science article on watershed management may draw on hydrology, ecotoxicology, environmental law, territorial planning, and the social sciences — and be relevant to researchers in

Table 1 – General comparison of the metrics discussed in this article

Indicator	Database	Window	Normalizes by field?	Weights prestige?	Level
JIF (JCR)	WoS	2 years	No	No	Journal
SNIP	Scopus	3 years	Yes	No	Journal
SJR	Scopus	3 years	Partially	Yes	Journal
CiteScore	Scopus	4 years	No	No	Journal
FWCI	Scopus / SciVal	Publication year + 3 years	Yes (field, year, type)	No	Article (aggregable)
Percentile/Quartile (JCR)	WoS	2 years (JIF-based)	Yes (by category)	No	Journal
Percentile/Quartile (Scopus)	Scopus	4 years (CiteScore-based)	Yes (by category)	No	Journal
Altmetrics (AAS)	Multiple digital sources	Real time	No	No	Article
PlumX Metrics	Scopus / Plum Analytics	Cumulative	No	No	Article
FCR	Dimensions	Cumulative since publication (minimum 2 years)	Yes (Field of Research)	No	Article (aggregable)
H5	Google Scholar / OpenAlex / Dimensions	5 full calendar years	No	No	Journal

all these areas. This dispersion of citations across different fields does not reduce the real impact of the work. However, because no specific field "adopts" the journal as its central reference, JIF values become diluted, capturing only a fraction of the actual influence.

How fields of knowledge are defined

The field-normalized indicators discussed so far — the SNIP, the FWCI, and the percentiles of the CiteScore and of the JCR — all depend on a prior subject classification of journals and articles: the 334 ASJC categories of Scopus and the equivalent WoS categories, already mentioned. This classification, however, rarely receives the same scrutiny devoted to the indicators that depend on it. The disadvantage faced by multidisciplinary fields in the metrics discussed here is an observable phenomenon regardless of how these categories are constructed — which is precisely why indicators such as the SNIP (Moed, 2010) and the FWCI (Waltman and van Eck, 2013) were conceived. Understanding how the categories are constructed does not call this finding into question; on the contrary, it helps explain why this disadvantage occurs with the observed intensity and where it could be mitigated. Below, we describe four approaches used to construct such categories: three widely established in the practice of indexing databases and a fourth, which appears as a comparative horizon, based on machine learning.

- Classification by journal — all articles published in the same venue inherit the subject category assigned to the journal as a whole. This is the most transparent and auditable scheme: any researcher can publicly consult the categories that Scopus or the WoS assigns to a given journal. Its weakness lies in the premise of thematic homogeneity within the journal — not always true, especially in multidisciplinary venues (Waltman and van Eck, 2012a) — and in its reliance on each database's editorial criteria. Wang and Waltman (2016), in a large-scale analysis, found that the WoS is more accurate than Scopus in journal-level classification, but that both databases contain a considerable number of journals with inaccurate classifications, indicating that this scheme is neither neutral nor infallible. This is the classification that underpins the normalization of the FWCI: although the assignment occurs at the article level, with articles fractionally assigned across categories, the categories themselves are inherited from the journal.
- Classification by publication — instead of inheriting the category of the journal, each article is classified based on its own characteristics, such as the predominant fields among the works it cites or that cite it. Klavans and Boyack (2017) compared three variants of this approach — direct citation, bibliographic coupling, and co-citation — and found that direct citation generates more accurate taxonomies, provided that the article has a sufficiently extensive reference list. This is a finer-grained classification than the journal-level one, but it is computationally demanding and potentially less accurate for documents with fewer references — such as short editorials, technical notes, or articles from fields with more parsimonious citation practices — given that the gold standard used by those authors was built with articles having at least one hundred references.
- Bibliometric networks (citation-based clustering) — instead of comparing articles individually, this method algorithmically groups the entire database into communities of documents

densely connected by citations, assigning the category directly to the document, without going through the journal. Waltman and van Eck (2012a) proposed the seminal methodology for this type of classification, now employed, with variations, both by CWTS/Leiden and by the commercial databases themselves. Because it does not depend on human editorial criteria, the method tends, in principle, to reduce the subjectivity of journal-level classification. It has, however, documented costs: Clarivate describes the algorithm employed by the WoS only in general terms, without publishing the full technical methodology (Clarivate, 2021b), which hinders the end user's understanding; and the Scopus implementation covers only 95% of the content indexed in the database, not its entirety (Elsevier, 2024). Two well-known implementations illustrate the approach, with structures that differ from each other: the Topics (grouped into Topic Clusters) of Scopus/SciVal, built through direct citation analysis (Elsevier, 2024); and the Citation Topics of the WoS, built by a community detection algorithm (Leiden) in a three-level hierarchy — micro, meso, and macro (Clarivate, 2021b).

- Classification by machine learning — still a comparative horizon: unlike the three previous schemes, already operational in the consolidated commercial databases, machine-learning classification appears in a more mature form on more recent platforms, such as OpenAlex and Dimensions/ANZSRC. Thelwall and Jiang (2025) conducted an independent study to precisely address this question, assessing whether OpenAlex is suitable for research evaluation. OpenAlex, for example, trains a language model — a variant of BERT — to classify each article on the basis of its title, abstract, citations, and journal name, achieving about 72% accuracy for the main category when all of this information is available; however, only half of the articles in the database have a usable abstract, which reduces accuracy precisely where it would be most needed (OpenAlex, 2024). This is the pattern of the method: it preserves granularity and reduces the decisive, exclusive weight of the journal — which becomes just one among several signals used by the model — but its quality is conditioned on the coverage of the training set and the richness of the metadata, penalizing fields with few labeled examples or with output for which no abstract is available.

It is worth relating this overview to the methodological choice described in Section "The Brazilian context: CAPES, the end of the Qualis list, and the new logic of article classification": in Procedure 1, CAPES (2025) classifies journals by the ASJC categories of Scopus and the equivalent WoS categories, assigning the article's stratum according to the journal's percentile within its own subject category — a public, stable scheme, auditable by any researcher, which has the additional advantage of neutralizing, by construction, the difference in citation density across fields discussed throughout this article. The vulnerability that persists is of another kind: as shown in this subsection, the very category to which a journal is assigned does not always accurately reflect its real thematic boundaries (Wang and Waltman, 2016) — so that the percentile of a multidisciplinary journal, calculated within a single category perhaps poorly suited to its scope, may not reflect its true position among comparable peers, even though the databases used already offer the citation-network alternative described above.

Time windows and distinct citation rhythms

In the applied social sciences, a book published today may be the central reference for a generation of researchers — but books do not enter into the calculation of the JIF. In the environmental sciences, long-term studies frequently accumulate citations over decades. In fields whose literature is predominantly in languages other than English, this underestimation of impact is even greater, since international indexing databases cover output in Portuguese, Spanish, Arabic, or Chinese unevenly.

In fields with smaller communities, indicators with wider windows — such as the SJR (3 years) and the CiteScore (4 years) — offer more representative readings, although they are still not ideal. Percentiles and quartiles are the most suitable instruments for these fields: by positioning each journal relative to its peers within the same category, they eliminate the scale bias that systematically penalizes such journals (Hicks et al., 2015; Waltman, 2016).

It is important to recognize that these limitations stem not only from the design of the indicators but also from the coverage of the databases that feed them. Scopus and the WoS favor journals indexed according to editorial criteria that are predominantly aligned with English-language publishing practices, which systematically underrepresent the output of Latin America, Africa, and Asia, as well as regional journals relevant to local public policy. The two databases also differ significantly in coverage — a journal may be indexed in one but absent from the other — so the choice of database is not neutral and can change the percentile assigned to the same article. Any limitation attributed to a bibliometric indicator is therefore also a limitation inherited from the database from which it is calculated.

Indicators as part of the promotion of scientific integrity

Paradoxically, fields with smaller communities are also the most vulnerable to the proliferation of predatory journals, a term used for journals whose editorial practices do not ensure the integrity of the publication process, as set out in the principles published by the Committee on Publication Ethics[8] (COPE) at the international level or by the Brazilian Association of Scientific Editors[9] (ABEC) at the national level. With fewer institutional resources and greater pressure to publish in indexed journals, researchers in the multidisciplinary

[8] https://publicationethics.org/
[9] https://www.abecbrasil.org.br

sciences, the applied social sciences, and the humanities are frequent targets of journals that simulate legitimacy without actually maintaining it (Grudniewicz et al., 2019). A study of 3,067 respondents from Brazilian graduate programs confirms this pattern: the decision to submit an article is rarely made in isolation, and the influence of advisors and peers can either protect against submission to predatory journals or, when uninformed, propagate it — so that researchers in training with less access to qualified mentoring tend to be more exposed to these practices (Carmo et al., 2026).

It is precisely in this context that bibliometric indicators fulfill their most important role: indexing in the JCR, Scopus, or other selective databases functions as a barrier to legitimacy — auditable, transparent, and continuously monitored. The indicators offer researchers a verifiable reference: even though the indicators are partial abstractions of reality, as Fernandes (2022) puts it, it is better to have them than not to have them — they are more reliable than the absence of any objective criterion when researchers select the journals in which their articles will be published. Nevertheless, the main reference for publication should be the seriousness and selectivity of the editorial policy and the forum in which the subject the publication addresses is debated, combined with a rigorous process of article review.

The Brazilian context: CAPES, the end of the Qualis list, and the new logic of article classification

For Brazilian researchers, understanding international bibliometric indicators has acquired a practical and urgent dimension as of 2025: the Coordination for the Improvement of Higher Education Personnel (CAPES) no longer publishes the list of journals with their respective Qualis classification — the so-called "Qualis list" (Carvalho et al., 2026). In this regard, it is important to differentiate between "Qualis" and the "Qualis list".

Since the 1990s, the CAPES evaluation areas have used an article classification methodology. Under this methodology, articles produced by *stricto sensu* graduate programs in Brazil in a given evaluation period (a triennium or quadrennium) were classified into different *strata* — in the 2013–2016 quadrennium, into eight *strata* (A1, A2, B1, B2, B3, B4, B5, and C); from the 2017–2020 quadrennium onward, into nine strata (A1, A2, A3, A4, B1, B2, B3, B4, and C). To perform this classification, the evaluation areas relied on the reputation of the venues where these articles were published, using the presumed quality of the articles, gauged by the quality of the journals. This methodology is "Qualis."

As a way of lending transparency to the evaluation process, CAPES released a list of journals and the classification used for these venues in the specific period (triennium or quadrennium). This is the "Qualis list", which became, *de facto*, "the Qualis" for the Brazilian scientific community. It is important to stress that this list was never complete: if no member of a Brazilian *stricto sensu* graduate program had published an article in each venue, that venue would not be included in the "Qualis list." Moreover, owing to its characteristics and limitations, this classification only makes sense within the scope of evaluating graduate programs, and it is entirely inappropriate to use it, as institutions do, to evaluate researchers in competitive examinations and selection processes of a different nature. The venue, however, would not thereby cease to be relevant. For example, if no researcher from a graduate program in computing had published in Nature, this would not mean that the journal had ceased to be important for the field, but it would not appear on the field's "Qualis list." For this reason, it is important to reinforce that "the Qualis" is not a process for evaluating journals. This mistaken perception on the part of the community, and especially of editors, created difficulties for CAPES and for the evaluation areas in releasing the "Qualis list."

Although it played an important role in introducing a culture of evaluation of scientific output in the country, the Qualis accumulated sustained criticism over the years: in some evaluation periods, the strata varied across areas, making comparison between fields impossible; its specific purpose as an a posteriori indicator of output was misunderstood; and it was used for purposes for which it was not designed, such as competitive examinations, among others. In addition, the system resulted in strategic publication behaviors oriented toward classification, rather than toward quality or dialogue with the respective communities.

From the 2017–2020 quadrennial evaluation cycle onward, most CAPES areas began adopting the journal percentiles from WoS and Scopus to directly evaluate articles produced by *stricto sensu* graduate programs in the country. Under this logic, a journal ranked in the 90th percentile of the WoS or Scopus in its subject category would assign the article the highest stratum — in this sense, the criterion became verifiable and updated annually. Some areas also began to incorporate the FWCI as a complementary indicator, capturing not only the quality of the publication venues but also the normalized impact of the output actually generated by the programs — a significant conceptual shift that reinforces the focus of evaluation on the article rather than on the journal. For areas where it was not possible to identify percentiles in WoS or Scopus, a methodology was used to aggregate national journals and determine percentiles, ensuring comparability across areas. In this methodology, the Google Scholar H5 was used as the bibliometric indicator of the journals.

This transition had direct implications for Brazilian researchers, program coordinators, and journal editors. For researchers, it means that knowing the percentile rank of the WoS or Scopus journal in which they intend to publish is now more relevant than knowing a journal's *stratum* on the "Qualis list." For program coordinators, it entails learning to report and interpret indicators that were previously intermediated by CAPES but that now need to be accessed directly on the JCR, Scopus, and SciVal platforms. For editors of internationally indexed Brazilian journals, understanding the percentile

rank of the journal within its category is fundamental for both editorial management and communication with authors and institutions.

As the article classification methodology advances, it becomes increasingly clear that the evaluation of graduate education should focus on the article rather than the journal. The CAPES reference document (CAPES, 2025) for the 2025–2028 quadrennium details three article classification procedures that may be adopted by the areas, individually or in combination.

Procedure 1 classifies the article on the basis of the bibliometric indicators of the journal in which it was published — an indirect analysis — using the percentiles of the CiteScore (Scopus) and of the JCR Impact Factor (Clarivate/WoS) or, alternatively, percentiles calculated on the basis of the H5 index[10] obtained via OpenAlex, Google Scholar or, as is the case for the Environmental Sciences Area, Dimensions, with due quality curation of the editorial processes by the Evaluation Areas. Journals are ranked within subject groupings — 334 ASJC (All Science Journal Classification) categories in Scopus and 236 in the WoS — and each article will be classified according to the highest percentile among the categories to which the journal belongs. The stratification of the article will result in eight levels (A1 to A8), distributed in classes 12.5 percentage points wide: articles published in journals with a percentile ≥ 87.5 will be classified in stratum A1, and articles published in journals with a percentile < 12.5 will be classified in stratum A8. Articles published in journals lacking any of the indicators provided for, or in journals that do not follow the good editorial practices of the Committee on Publication Ethics[11] (COPE), will be classified in stratum C.

It is worth noting a limitation of the h-index — of which the H5 is a variant — demonstrated by Waltman and van Eck (2012b), one that can generate inconsistent rankings: two researchers or journals whose performance improves by exactly the same magnitude may nevertheless have their relative positions inverted, a result that the authors themselves consider difficult to justify as a measure of scientific impact. For this reason, the CWTS, the center where much of the normalization methodology discussed in this article originated, does not adopt it as an indicator in its rankings, such as the Leiden Ranking. This fragility warrants caution when interpreting marginal differences between strata calculated from this indicator, whose use is justifiable only in a complementary way and only when there is truly no adequate coverage in the international databases.

Procedure 2 starts from the result of Procedure 1 and adds further layers of analysis: qualitative criteria for the journal — which may include the valorization of national journals, open access, and indexing in databases relevant to the area — and direct bibliometric indicators of the article, such as the number of citations received in the Scopus, WoS, OpenAlex, Altmetrics, Dimensions, and CrossRef databases.

Procedure 3, in turn, consists of a qualitative analysis of the article, applied only to the best products of the program — given the level of detail required, it is not feasible to apply it to the entire output — and assesses aspects such as the relevance of the topic, conceptual advancement, and scientific contribution, resulting in a five-level scale (Very Good, Good, Fair, Weak, and Insufficient).

This methodology places the indicators discussed in this article — percentiles, CiteScore, Impact Factor, FWCI, and Altmetrics — in an operationally relevant position in the evaluation of Brazilian graduate education, making mastery of these indicators an essential competence for researchers and program coordinators.

The change at CAPES, discussed and approved by the Technical-Scientific Council for Higher Education (CTC-ES)[12], ultimately represents a convergence of the Brazilian evaluation system with international standards — which makes understanding the bibliometric indicators presented in this article not only academically relevant but also operationally necessary for any researcher involved in Brazilian graduate education. It is important to stress that the point is not to suggest that researchers and graduate students apply their understanding of the metrics strictly, but rather that they understand them to make better decisions in the context of disseminating their research results and understanding the evaluation process conducted by CAPES.

The responsible assessment movement

The recognition that bibliometric indicators are necessary but insufficient has, in recent decades, given rise to an international movement for responsible research assessment, which is today one of the central topics of metascience. This movement does not reject metrics; it proposes balancing them with qualitative peer judgment and fitting them to the purpose and context of each evaluation. Its two most influential milestones are the San Francisco Declaration on Research Assessment (DORA, 2013), which recommends not using the journal Impact Factor to evaluate individual researchers or articles, and the Leiden Manifesto (Hicks et al., 2015), which synthesizes ten principles for the use of indicators — among them, that metrics should support, and not replace, qualitative assessment, and that indicators must be appropriate to their purpose and sensitive to the particularities of each field. It is worth noting that DORA's recommendation regarding the use of the JIF in the evaluation of individual articles is not consensual: Waltman and Traag (2021) argue that this practice is not necessarily incorrect from a statistical standpoint, provided that the skewness of the journal's citation distribution is duly considered

[10] The H5 (h5-index) is the h-index calculated over the articles published in the last five complete calendar years of a journal or venue, according to the Google Scholar Metrics methodology: a journal has h5 = N if N of its articles published in that period received at least N citations each, and the remaining articles received N citations or fewer.

[11] https://publicationethics.org/

[12] The composition of the CTC is detailed at https://www.gov.br/capes/pt-br/acesso-a-informacao/participacao-social/conselhos-e-orgaos-colegiados/conselho-tecnico-cientifico-da-educacao-superior/composicao

in the interpretation — which keeps alive the discussion about the exact limits of this recommendation.

In the same direction, the Metric Tide report (Wilsdon et al., 2015) consolidated the notion of responsible metrics along five dimensions: robustness, humility, transparency, diversity, and reflexivity. These dimensions translate, in operational terms, the idea that runs through this article: an indicator is only informative when its limits are known, when it is applied to the field for which it was designed, and when it is interpreted considering the different citation culture. The diversity dimension, in particular, speaks directly to the situation of fields with smaller communities by recommending that evaluation encompass the plurality of areas, languages, and output formats, avoiding the systematic penalization of what does not conform to high-citation-density areas.

Along the same lines, Barra and Rojas-Hernandez (2022) argue that academic evaluation criteria should be broadened beyond bibliometric indicators, also valuing work with communities and applied interdisciplinary research aimed at understanding socio-environmental problems and at proposing solutions that improve the population's quality of life and protect nature.

More recently, the Agreement on Reforming Research Assessment and the Coalition for Advancing Research Assessment (CoARA, 2022) broadened this consensus by bringing together hundreds of institutions around a common commitment: to base evaluation primarily on qualitative judgment, supported by the responsible use of quantitative indicators. This principle aligns with the Brazilian transition discussed here, in which the CAPES article classification combines bibliometric indicators of journals and articles, qualitative aspects of journals (Procedures 1 and 2), and the qualitative analysis of the best products (Procedure 3). In short, the responsible assessment movement does not set metrics against quality: it reaffirms that indicators are legitimate instruments when used as an initial reference and with awareness of their limits — exactly like a thermometer, accurate for measuring temperature but unsuitable for measuring distances.

Final considerations

With each release of bibliometric indicators by Clarivate or Scopus, for example, the scientific community is confronted with a substantial volume of new numbers that will be immediately incorporated into institutional evaluations and will serve as a guide for researchers in choosing where to publish. In Brazil, the transition in the CAPES evaluation aims to make it clearer that part of the evaluation of programs focuses on appraising the quality of the articles, and not of the venues in which they are published, although these are important. The use of international percentiles and of other article-level indicators, such as the FWCI, makes understanding these instruments even more urgent. This periodic updating also implies that a journal's position — its percentile, quartile, or stratum — is not permanent: it can vary from one edition to the next even without any change in the journal's quality, especially in subject categories with a smaller number of journals, in which small oscillations in citation counts shift proportionally more positions in the ranking.

Bibliometric indicators are valuable and necessary tools, despite their limitations. They are not perfect, but they offer something that purely subjective evaluation cannot guarantee and that those being evaluated would not otherwise accept: objectivity, replicability, and comparability. In a scientific environment in which the proliferation of questionable publications poses a real threat to the integrity of knowledge, having reliable references to gauge the quality of articles using journal bibliometric indicators is not an academic luxury but a practical necessity. This aspect becomes even more important when the volume of articles is on the order of thousands.

The key to the appropriate use of these indicators lies in understanding their conceptual basis and their limits. The JIF of the JCR is robust within its own field, indicating the journal's reputation, but it has serious limitations when used as a proxy for gauging article quality. The SNIP and the SJR offer complementary perspectives in interdisciplinary contexts. The CiteScore widens the time window. The FWCI normalizes impact by field, year, and document type. Percentiles and quartiles translate performance into relative position. Altmetrics and PlumX broaden the horizon beyond the formal scientific literature. Each indicator answers a specific question; none of them answers all questions — and it is precisely for this reason that understanding them is indispensable to using them responsibly.

Finally, it is worth noting another important development to be analyzed in the future: the consolidation of open platforms such as Dimensions, OpenAlex, and Semantic Scholar, which represents one of the most significant transformations in contemporary bibliometrics. Unlike the WoS and Scopus, which have historically combined the production of indicators with selective processes of indexing and of editorial monitoring, these new initiatives adopt a predominantly inclusive logic, seeking to represent the universe of scientific communication in its greatest possible breadth. As a consequence, they have also begun to provide impact metrics, citation networks, and normalized indicators that, in many cases, are conceptually close to those produced by the traditional databases.

However, interpreting these indicators requires caution. In selective databases, indicators are calculated based on a set that has been subjected to criteria for editorial quality, publication regularity, process transparency, and scientific integrity. In broad-coverage platforms, by contrast, the quality of the indicators depends heavily on the data sources used, on the consistency of the metadata, and, above all, on the quality of the journals originally incorporated into the system. Thus, although they increase the visibility of areas, languages, and regions historically underrepresented in traditional scientific indexing systems, these new open-science platforms are still developing and consolidating, facing the challenge of strengthening journal in-

dexing processes tied to editorial rigorous screening, as those adopted by consolidated databases.

This distinction reveals a fundamental question for the future of research evaluation. While selective systems tend to offer greater control over the quality of the universe analyzed, open platforms provide greater coverage, transparency, and access to scientific data. The trend observed in recent years suggests that these models should not be understood as competitors, but as complementary. The former continues to play an important role in preserving editorial integrity and in qualifying the indicators, while the latter expand the capacity to map global science in all its diversity. The challenge for researchers, evaluators, and science policymakers is now to understand the potential and the limitations of each approach simultaneously.

For the Brazilian scientific community, investing in understanding these instruments, both traditional and emerging, is investing in the capacity to conduct science with greater awareness, rigor, and intellectual autonomy.

Authors' Contributions: Fernandes, V.: Conceptualization, Formal Analysis, Methodology, Writing – original draft, Writing – review & editing; Zorzo, A.F.: Formal Analysis, Writing – review & editing; Souza Filho, A.G.: Supervision, Writing – review & editing.

Conflicts of interest: The authors declare no conflicts of interest.

Funding: none.

Declaration on the use of generative artificial intelligence: The authors declare that they used generative artificial intelligence tools to support the preparation of this article. Such tools were employed on the following specific fronts: (i) Claude (Anthropic), to assist in editing the figures that illustrate the text; (ii) Grammarly Desktop, for the review of grammar, spelling, and clarity. No scientific content — data, analyses, results, interpretations, or conclusions — was produced autonomously by these tools. The authors assume full responsibility for the integrity, accuracy, and originality of the work.

References

Barra, R.O.; Rojas-Hernandez, J., 2022. The indexing of scientific knowledge: The need for knowledge at the service of community development and nature protection. Revista Brasileira de Ciências Ambientais, v. 57 (4), 689-692. https://doi.org/10.5327/Z217694781470

Bergstrom, C., 2007. Eigenfactor: Measuring the value and prestige of scholarly journals. College & Research Libraries News, v. 68 (5), 314-316. https://doi.org/10.5860/crln.68.5.7804

Bornmann, L., 2014. Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. Journal of Informetrics, v. 8 (4), 895-903. https://doi.org/10.1016/j.joi.2014.09.005

Brin, S.; Page, L., 1998. The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems, v. 30 (1-7), 107-117. https://doi.org/10.1016/S0169-7552(98)00110-X

Carmo, F.A.; Rezende, S. O.; Paxiúba, C. M. C.; Lobato, F. M. F., 2026. Predatory publishing in the academic landscape: researchers' awareness and vulnerabilities. Scientometrics, v. 131, 3029-3054. https://doi.org/10.1007/s11192-026-05614-0

Carvalho, D.; Souza Filho, A. G.; Sarmento, N., 2026. CAPES e o novo sistema de classificação de artigos científicos: o que é verdade e o que é desinformação. https://theconversation.com/capes-e-o-novo-sistema-de-classificacao-de-artigos-cientificos-o-que-e-verdade-e-o-que-e-desinformacao-285209

Clarivate, 2021a. Introducing the Journal Citation Indicator: A new, field-normalized measurement of journal citation impact. Clarivate. https://clarivate.com/blog/introducing-the-journal-citation-indicator-a-new-field-normalized-measurement-of-journal-citation-impact/

Clarivate, 2021b. Introducing Citation Topics. Clarivate. https://clarivate.com/academia-government/blog/introducing-citation-topics/

Clarivate, 2024. Category Normalized Citation Impact (CNCI). InCites Indicators Handbook. https://incites.help.clarivate.com/Content/Indicators-Handbook/ih-normalized-indicators.htm

Coalition for Advancing Research Assessment (CoARA), 2022. Agreement on Reforming Research Assessment. https://zenodo.org/records/13480728.

Colledge, L.; Moya-Anegón, F.; Guerrero-Bote, V.; López-Illescas, C.; El Aisati, M. et al. 2010. SJR and SNIP: two new journal metrics in Elsevier's Scopus. Serials: The Journal for the Serials Community, v. 23 (3), 215-221. https://serials.uksg.org/articles/10.1629/23215

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), 2025. Documento Referencial: Diretrizes comuns da Avaliação de Permanência dos Programas de Pós-Graduação stricto sensu — ciclo avaliativo 2025–2028. CAPES, Brasília. https://www.gov.br/capes/pt-br/centrais-de-conteudo/documentos/avaliacao/19052025_20250502_DocumentoReferencial_FICHA.pdf.

DORA, 2013. San Francisco Declaration on Research Assessment. https://sfdora.org

Elsevier, 2019. Research Metrics Guidebook. Elsevier B.V., March 2019. https://assets.ctfassets.net/o78em1y1w4i4/54o8UH5cdisCqfUFpSG03T/b1c80b9c4a9fcf168f85065a0f044a64/ELSV-13013-Elsevier-Research-Metrics-Book-r12-WEB.pdf

Elsevier, 2024. SciVal Topics [technical documentation]. https://service.elsevier.com/app/answers/detail/a_id/27947/supporthub/scopus/kw/topics/

Eysenbach, G., 2011. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. Journal of Medical Internet Research, v. 13 (4), e123. https://doi.org/10.2196/jmir.2012

Fernandes, V., 2022. Why and where to publish. Revista Brasileira de Ciências Ambientais, v. 57 (3), 516-518. https://doi.org/10.5327/Z2176-94781439

Garfield, E., 1972. Citation analysis as a tool in journal evaluation. Science, v. 178 (4060), 471-479. https://doi.org/10.1126/science.178.4060.471

Garfield, E., 2006. The history and meaning of the journal impact factor. Journal of the American Medical Association, v. 295 (1), 90-93. https://doi.org/10.1001/jama.295.1.90

González-Pereira, B.; Guerrero-Bote, V.P.; Moya-Anegón, F., 2010. A new approach to the metric of journals' scientific prestige: The SJR indicator. Journal of Informetrics, v. 4 (3), 379-391. https://doi.org/10.1016/j.joi.2010.03.002

Guerrero-Bote, V.P.; Moya-Anegón, F., 2012. A further step forward in measuring journals' scientific prestige: The SJR2 indicator. Journal of Informetrics, v. 6 (4), 674-688. https://doi.org/10.1016/j.joi.2012.07.001

Grudniewicz, A.; Moher, D.; Cobey, K.D.; Bryson, G.L.; Cukier, S. et al. 2019.

Predatory journals: no definition, no defence. Nature, v. 576, 210-212. https://doi.org/10.1038/d41586-019-03759-y

Hicks, D.; Wouters, P.; Waltman, L.; Rijcke, S.; Rafols, I., 2015. Bibliometrics: The Leiden Manifesto for research metrics. Nature, v. 520 (7548), 429-431. https://doi.org/10.1038/520429a

Klavans, R.; Boyack, K.W., 2017. Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? Journal of the Association for Information Science and Technology, v. 68 (4), 984-998. https://doi.org/10.1002/asi.23734

Larivière, V.; Kiermer, V.; MacCallum, C. J.; McNutt, M.; Patterson, M. et al. 2016. A simple proposal for the publication of journal citation distributions [preprint]. bioRxiv. https://doi.org/10.1101/062109

Larivière, V.; Sugimoto, C. R., 2019. The Journal Impact Factor: A brief history, critique, and discussion of adverse effects. In: Glänzel, W.; Moed, H.F.; Schmoch, U.; Thelwall, M. (Eds.), Springer Handbook of Science and Technology Indicators. Springer, Cham, pp. 3-24. https://doi.org/10.1007/978-3-030-02511-3_1

Leydesdorff, L.; Rafols, I., 2011. Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. Journal of Informetrics, v. 5 (1), 87-100. https://doi.org/10.1016/j.joi.2010.09.002

Moed, H.F., 2010. Measuring contextual citation impact of scientific journals. Journal of Informetrics, v. 4 (3), 265-277. https://doi.org/10.1016/j.joi.2010.01.002

Moed, H.F.; van Leeuwen, T.N.; Reedijk, J., 1998. A new classification system to describe the ageing of scientific journals and their impact factors. Journal of Documentation, v. 54 (4), 387-419. https://doi.org/10.1108/EUM0000000007175

Oliveira Jr., E.; Zorzo, A. F., 2026. Rethinking Research in Computer Science: Beyond Metrics, Toward Scientific Evolution. IEEE Computer. https://doi.org/10.1109/MC.2026.3680012

OpenAlex, 2024. Topics [technical documentation]. https://help.openalex.org/hc/en-us/articles/24736129405719-Topics

Page, L.; Brin, S.; Motwani, R.; Winograd, T., 1999. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report, Stanford InfoLab. https://homepages.dcc.ufmg.br/~nivio/cursos/ri11/sources/pagerank.pdf

Priem, J.; Taraborelli, D.; Groth, P.; Neylon, C., 2010. Altmetrics: A manifesto. http://altmetrics.org/manifesto

Rafols, I.; Leydesdorff, L.; O'Hare, A.; Nightingale, P.; Stirling, A., 2012. How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management. Research Policy, v. 41 (7), 1262-1282. https://doi.org/10.1016/j.respol.2012.03.015

SciVal, 2026. Research bibliometric analysis platform. https://www.scival.com.

Seglen, P.O., 1997. Why the impact factor of journals should not be used for evaluating research. British Medical Journal, v. 314 (7079), 498-502. https://doi.org/10.1136/bmj.314.7079.497

Thelwall, M.; Fairclough, R., 2015. Geometric journal impact factors correcting for individual highly cited articles. Journal of Informetrics, v. 9 (2), 263-272. https://doi.org/10.1016/j.joi.2015.02.004

Thelwall, M.; Haustein, S.; Larivière, V.; Sugimoto, C.R., 2013. Do altmetrics work? Twitter and ten other social web services. PLOS ONE, v. 8 (5), e64841. https://doi.org/10.1371/journal.pone.0064841

Thelwall, M.; Jiang, X., 2025. Is OpenAlex suitable for research quality evaluation and which citation indicator is best? Journal of the Association for Information Science and Technology, v. 76 (12), 1660-1681. https://doi.org/10.1002/asi.70020

Waltman, L., 2016. A review of the literature on citation impact indicators. Journal of Informetrics, v. 10 (2), 365-391. https://doi.org/10.1016/j.joi.2016.02.007

Waltman, L.; Traag, V. A., 2021. Use of the journal impact factor for assessing individual articles: Statistically flawed or not? F1000Research, v. 9, 366. https://doi.org/10.12688/f1000research.23418.2

Waltman, L.; van Eck, N.J., 2012a. A new methodology for constructing a publication-level classification system of science. Journal of the American Society for Information Science and Technology, v. 63 (12), 2378-2392. https://doi.org/10.1002/asi.22748

Waltman, L.; van Eck, N.J., 2012b. The inconsistency of the h-index. Journal of the American Society for Information Science and Technology, v. 63 (2), 406-415. https://doi.org/10.1002/asi.21678

Waltman, L.; van Eck, N.J., 2013. A systematic empirical comparison of different approaches for normalizing citation impact indicators. Journal of Informetrics, v. 7 (4), 833-849. https://doi.org/10.1016/j.joi.2013.08.002

Waltman, L.; van Eck, N.J.; van Leeuwen, T.N.; Visser, M.S., 2013. Some modifications to the SNIP journal impact indicator. Journal of Informetrics, v. 7 (2), 272-285. https://doi.org/10.1016/j.joi.2012.11.011

Wang, Q.; Waltman, L., 2016. Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. Journal of Informetrics, v. 10 (2), 347-364. https://doi.org/10.1016/j.joi.2016.02.003

Wilsdon, J.; Allen, L.; Belfiore, E.; Campbell, P.; Curry, S. et al. 2015. The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. https://www.ukri.org/wp-content/uploads/2021/12/RE-151221-TheMetricTideFullReport2015.pdf