

Ethics and bias in research metrics: A comprehensive review of challenges, manifestations, and pathways to reform

Rouzbeh Ghousi*

^aDepartment of Industrial Engineering, School of Engineering, Iran University of Science and Technology, Tehran, Iran

CHRONICLE

Article history:

Received: October 1, 2025

Received in revised format: October 7, 2025

Accepted: October 7, 2025

Available online:

November 7, 2025

Keywords:

Research Metrics

Ethics

Bias

Responsible Research Assessment

Scientometrics

Research Integrity

Gaming

Gender Bias

Geographical Bias

ABSTRACT

The widespread application of quantitative measurements in evaluating research, though very attractive for its supposed objectivity and expediency, has on the other hand given rise to an intricate web of ethical issues and biases in the system. The current review not only critiques but strategically moves through a thorough system analysis revealing the limitations of metrics and their sociotechnical implications. Our first step is to map out the ethics involved and thereby set up rules for the proper use of metrics. The next stage is to look into the bias aspect of metrics and the various forms of bias such as issues of location and language, unfairness among different fields, and the ongoing divide between the genders. The whole matter of metric malpractice—gaming, manipulation, and the detrimental over-optimization of research integrity—are some of the things that we have extensively discussed in this paper. Likewise, we have raised the emerging trend's ethical implications, namely, altmetrics, visualization, and algorithmic evaluation, taking into account their capability of both widening influence and introducing additional types of bias. Alongside this, we provide a picture of recent empirical findings about the status of research ethics and the level of support from institutions. Lastly, we bring together a progressive agenda for change, which includes institutional accountability, the shaping of reflexive evaluation methodologies, and the essential incorporation of qualitative, expert opinion. We express that a major change in mentality is necessary—one that will place metrics in a supportive role in a holistic, qualitative, and ethically-grounded research evaluation ecosystem.

© 2025 by the authors; licensee Growing Science, Canada.

1. Introduction

The quantification of scientific research has made its mark and one cannot overlook it in the modern academic scene. Numerical indicators like journal impact factors, citation counts, h-indices, and altmetrics all at once promise a quick, easy, and non-subjective way to evaluate the quality, impact, and productivity of researchers, institutions, and even countries. This phenomenon, which has been called "metric tide," is the result of the aforementioned pressures on research systems: there is increased competition for funding, demands for public accountability, and the sheer volume of scholarly output. On the other hand, the use of metrics has resulted in a lot of criticism from various quarters. These include scientists, sociologists of science, and research policymakers. Boasting the most correlative scientific arguments, this paper ranks the first ones: research metrics application is neither neutral nor value-free. It is rather the case that the whole process is heavily laden with ethical dilemmas and is subject to various kinds of biases that might affect the entire scientific enterprise, thereby facilitating inequities and undermining the very basis of research trust. The simple number's attraction often makes it difficult to see the intricate and oftentimes dependent on the situation nature of scientific progress, hence the birth of "Goodhart's Law," which articulates that when a measure is set as a target, it loses its good measurement quality.

* Corresponding author.

E-mail address: ghousi@iust.ac.ir (R. Gousi)

Numerous ethical challenges confront the scientific community. These go all the way from overt misbehavior, such as gaming and manipulating of indicators, to the subtler, systemic problems that include the very choice of metrics which, unintentionally, promote certain types of research, certain groups of researchers, and certain regions of the world over others. The use of metrics at the level of journals, such as the Impact Factor, for assessing the value of individual articles or researchers has been widely criticized as invalid and unethical, yet the practice continues to be very common. In addition to this, there are also new metrics based on social and online media—altmetrics—which present both the opportunity of getting a wider view of impact and the risk of manipulation and the preference for popularity over substance.

The first one will set the stage by describing the ethical frameworks and principles that have been suggested as the pillars for Responsible Research Assessment, like the San Francisco Declaration on Research Assessment (DORA) and the Leiden Manifesto. Then it will go on to carry out a systematic investigation of the most important categories of bias that are tied to bibliometric and altmetric data, which are geographical, linguistic, disciplinary, and gender biases. It will come to the fourth point where it will be talking about the metrics-related problems such as the assumption of the 'correct' order of publication, the arts of 'hiding' unproductive researchers, and the making of untrustworthy communities. Next, it will analyze the new frontiers 'algorithmic evaluation and visualization of science' from the ethical point of view. Lastly, the review will bring together the overlapping views on reform, highlighting the significance of institutions, funders, and publishers in the transition to a scenario where metrics are an accompanying factor to expert judgment rather than a substitute.

This review aims to illuminate a pivotal matter that lies at the intersection of modern science policy and communication through the integration of recent empirical findings, theoretical critiques, and practical guidelines. A more sophisticated, fair, and integrity-driven mindset towards the comprehension and evaluation of scientific inquiry is hoped to be the outcome of the study.

2. Foundations: Framing the Ethical Landscape of Research Assessment

The moral dilemmas of using metrics in research are not merely isolated incidents, they are rather a big part of the very structure and basic assumptions behind these systems. The metrics are often considered to be mirroring the very hidden biases—preferring certain fields of study, publication types, or areas of the world—thus influencing the academic actions in a way that could possibly drag down the values of equity and integrity. To engage critically with these topics, it is very important to trace the ethical frameworks that have come up to assist the responsible use of metrics. The building blocks of the frameworks are the principles of transparency, accountability, inclusivity, and contextual relevance. The Leiden Manifesto and the San Francisco Declaration on Research Assessment (DORA) are two examples of frameworks that disagree with using only simple quantitative indicators and instead promote the idea of more complex, qualitative evaluations. When ethical considerations take the front seat, the universities and the researchers can skillfully walk through the challenging landscape of performance measurement and get the metrics to function as tools for insight rather than instruments of distortion. The use of metrics in a responsible manner calls for constant reflection, conversation, and changes to the system.

2.1. *The Shift from Governance to Accountability*

The historical evolution of research evaluation marks a profound transformation in how scientific work is assessed and governed. Traditionally, the academic community relied on peer-based, qualitative judgment rooted in professional norms and Mertonian ideals such as communalism, universalism, and disinterestedness. These principles emphasized the intrinsic value of scientific inquiry and the collective advancement of knowledge. However, beginning in the late 20th century and accelerating into the 21st, academia experienced a paradigm shift toward “new public management.” This model introduced mechanisms of accountability, audit culture, and performance indicators, reshaping the landscape of research governance.

As Whitley (2020) observes, this transition shifted authority from scientific communities to administrators and policymakers, who increasingly prioritized quantifiable outputs—such as publication counts, citation metrics, and grant income—over qualitative dimensions like originality, methodological rigor, and long-term societal relevance. The rise of this metric-driven culture, often referred to as the “metric tide,” created incentives that could distort research behavior, encouraging strategic publishing and metric gaming. While metrics can offer useful insights, their dominance risks marginalizing nuanced, context-rich evaluations. Understanding this historical trajectory is essential for developing more balanced and ethically grounded approaches to research assessment—ones that honor both quantitative indicators and the deeper values of scientific inquiry.

2.2. *Core Ethical Principles and Manifestos*

In reaction to the increasing discomfort surrounding this quantification, the main principles for the ethical use of metrics have been expressed through a number of notable manifestos and reports.

The San Francisco Declaration on Research Assessment (DORA): The San Francisco Declaration on Research Assessment (DORA), first introduced in 2012 and still shaping global research policy, has been and still is the major movement that has ever had a large impact on the ethical reform of the research evaluation. With its challenging of the Journal Impact Factor (JIF) and critiquing of the latter's use as a proxy for research quality, DORA has already gotten to the point where it

was being treated in funding, hiring, and promotion decisions. The declaration insists on a total change in the criteria of assessing scholarly contributions, stressing that the evaluation of research must be based on its innate worth and not on the reputation of the journal in which it is published. It further recommends the use of metrics at the article level and the employment of various indicators which would reflect a much wider range of impact including societal relevance, openness, and engagement. DORA's ethical positioning is based on the notions of fairness, transparency, and validity, and its proponents maintain that applying journal-based measures to evaluate single researchers is not only a statistically invalid at least but also an unethical practice. DORA by advocating for more detailed and context-sensitive evaluation practices strives to shift the academic rewards systems to be in line with the values of scientific integrity and openness. Its influence has grown and allows the policies of institutions to be more aligned with the change and also promote more equitable and responsible research cultures throughout the world.

The Leiden Manifesto, as articulated by Hicks et al. (2022), offers a robust and pragmatic framework for ethical research evaluation. It distills best practices into a set of guiding principles that serve as a moral compass for institutions, funders, and evaluators. Central to the manifesto is the idea that quantitative metrics should complement—not replace—expert qualitative judgment. This foundational stance reasserts the role of peer review and contextual understanding in assessing scholarly work. The manifesto also emphasizes the importance of aligning evaluation with the specific missions of researchers and institutions, rejecting uniform standards that ignore disciplinary and regional diversity. It advocates for the recognition of locally relevant research, challenging systemic biases that privilege mainstream topics and elite institutions. Transparency and simplicity in data collection and analysis are highlighted as ethical imperatives, ensuring that evaluation processes remain accountable and comprehensible. Moreover, the right of researchers to verify the data used in their assessment is upheld as a matter of procedural fairness. Collectively, these principles form a comprehensive ethical architecture that resists the reductive and opaque use of metrics. The Leiden Manifesto encourages a more thoughtful, inclusive, and responsible approach to research evaluation—one that honors the complexity and diversity of scientific inquiry.

The Metric Tide Report, led by Wilsdon et al. (2021), marked a pivotal moment in the ethical discourse surrounding research evaluation. It introduced the concept of “responsible metrics,” defined as the responsible use of quantitative indicators in the governance, management, and assessment of research. This framework does not advocate for the abandonment of metrics, but rather for their thoughtful and context-sensitive application. The report outlines five core ethical principles—robustness, humility, transparency, diversity, and reflexivity—that serve as a foundation for evaluating the appropriateness and impact of metric use. These principles encourage evaluators to recognize the limitations of indicators, avoid over-reliance on simplistic measures, and remain attentive to disciplinary and institutional diversity. Wilsdon (2023) further emphasizes that responsible metrics are not merely technical tools but instruments of trust-building. They must reinforce, rather than undermine, the values of a healthy research culture. By promoting metric literacy and ethical awareness, the Metric Tide Report calls for a shift in institutional behavior—from metric-driven governance to reflective and inclusive evaluation practices. Its influence continues to shape policy frameworks and institutional guidelines, offering a balanced approach that integrates quantitative insight with qualitative judgment in the pursuit of research excellence.

2.3. A Synthesis of Ethical Imperatives

The foundational frameworks like the ten principles for responsible research assessment put forward by Moher and Naudet (2021) have been used as a basis from which ethical metrics for research evaluation have been established. This consensus represents the gradually growing awareness that merely counting publications and citations is a very simplistic way of looking at scholarly contributions. Holism demands a more encompassing perspective—one that takes into account diverse outputs, societal impacts, and research practices. Contextualization underlines the interpretation of metrics in light of the specific researcher's area of expertise, institutional mission, and career stage, therefore avoiding any blanket judgments. Transparency guarantees that the data, methods, and criteria relied on in evaluations are subjected to scrutiny, thereby creating a climate of accountability and trust. Integrity requires that evaluation systems not only discourage metric manipulation but also uphold the values of honest inquiry. Equity draws attention to the requirement for designing assessments that proactively recognize and counteract systemic biases, whether they are disciplinary, geographic, or demographic. Thus, the ethical principles are the guiding stars in the complicated process of dealing with metric-based assessment. This method is a one that commits to the very diversity of research ecosystems and, hence, admits to being more rigorous, fair, and inclusive in scholarly evaluation than the previous one.

3. The Anatomy of Bias in Research Metrics

One of the main ethical criticisms against research metrics is that they are not the objective and culture-neutral tools that they are mostly assumed to be. On the contrary, they are deeply rooted in and can even promote the social, economic, and epistemic biases that are already present in the worldwide research system.

3.1. Geographical and Linguistic Bias: The Perpetuation of a Global Divide

The scientific production and impact all over the world are very unevenly distributed and bibliometric indicators often reflect and even magnify this inequality instead of just recording it. According to Larivière, Haustein, and Mongeon (2020),

the scientific publishing industry has taken over globally, making commercial publishers the only ones in the driver's seat. These publishers are mainly from the Global North and make up the oligopoly that has surrounded the world of science. They run the most reputable journals, which gain the most citations and are most visible because of it. This situation imbalances the whole system of scholarly communication and forces the researchers from the Global South to go for publication in international English-language outlets if they want to get recognized. Nevertheless, Larivière and Sugimoto (2022) observe that this leads to a double bind: the output of local or regional journals is often unrecognized in global bibliometric databases; on the other hand, the output of international venues might lose its relevance for local contexts and needs. Monjeon and Paul-Hus (2022) broaden the discussion by arguing the case of ethics in database coverage and showing that the platforms like Web of Science and Scopus selectively index journals, systematically excluding many from the Global South. The exclusion distorts the global map of science thereby reinforcing inequities in visibility, impact and resource allocation. Only by changing how research metrics are constructed and applied to address these biases will it become possible to do so; greater emphasis will need to be on inclusivity, diversity, and the ethical implications of coverage and representation.

Linguistic bias in research metrics is a pervasive issue that closely parallels geographical disparities in scholarly visibility. Citation databases are overwhelmingly skewed toward English-language publications, creating systemic disadvantages for researchers who publish in other languages such as French, Mandarin, or Swahili. As a result, their work often receives fewer international citations—not due to its quality, but because of linguistic barriers that hinder discovery and uptake. This imbalance raises a profound ethical concern. Costas and van Eck (2022), in their work on ethical visualization, argue that scientific maps must not conflate database coverage with actual global scientific activity. When they do, entire linguistic traditions and regions risk being rendered invisible, distorting our understanding of global knowledge production. Adjovi (2025) reinforces this critique through a comprehensive bibliometric analysis, showing that even the norms and discourse surrounding research ethics are frequently shaped by Anglo-American perspectives. This dominance influences how ethical research is defined, assessed, and rewarded worldwide. To address these inequities, there is a growing call to decolonize the frameworks underpinning research metrics. Only by embracing linguistic and cultural diversity can ethical evaluation systems become truly inclusive, ensuring that all scholarly voices—regardless of language or location—are recognized and valued in the global research ecosystem.

3.2. *Disciplinary Bias: The Illusion of a Level Playing Field*

There exists an issue in bibliometrics that is both widespread and raises serious ethical questions, which is the bias mainly resulting from the different citation practices among academic disciplines. The comparison of the number of citations of a mathematician, a historian, and a molecular biologist would be considered a very serious category error, which would then be reflected in valuation of knowledge and in the assessing of interdisciplinary collaboration. As pointed out by Bornmann (2021), normalization is both a methodological necessity and an ethical imperative. If the citation rates are not adjusted according to the field and the publication year, the counts are misleading. Fast publication cycles, extensive co-authorship, and cultures of dense referencing are characteristics of disciplines like biomedicine and astrophysics and lead to high citation volumes. On the other hand, fields such as mathematics or the humanities generally have low citation volumes because they produce single-author publications that adhere to different citation norms and hence have lower raw counts. If metrics are applied uncritically in the making of funding or promotion decisions, those disciplines that produce lots of citations would automatically be favored while the less active ones would be deprived of their valuable contributions. Rafols and Yegros-Yegros (2020), along with Rafols (2021), further reveal that interdisciplinary research suffers a “citation penalty” due to its cross-cutting nature, which limits its visibility within any single disciplinary silo. This bias discourages the kind of integrative research needed to solve complex societal problems. Ethical evaluation requires that the metrics are normalized and provided with context in a manner that maintains and even promotes the diversity of thought and the corresponding innovations.

3.3. *Gender Bias: Uncovering Systemic Inequities*

An increasing number of studies has indicated that scientific metrics are not at all gender-neutral and, in many cases, strengthen the unfairness that women and other minority groups face in science. Didegah and Thelwall (2021) have demonstrated that, after taking field and other factors into account, men-led articles tend to get more citations than women-led ones. This citation gap that has not gone away is a result of many interacting factors like the unequal access to collaborators, the difference in promotion practices, and the subconscious bias in the citations made. The Matthew Effect—wherein the established researchers amass unfairly huge recognition—adds to these inequities and frequently makes the male researchers the beneficiaries. Paul-Hus and Desrochers (2021) draw attention to how these biases get into citation networks, thus maintaining the structural imbalances. Sugimoto (2023) voices the same concern regarding the reliance on citation totals but in a more critical way; she claims that the conventional metrics are too narrow to spot the major academic contributions such as teaching, coaching, and public engagement—areas where female and minority scholars are usually more involved. Even the metrics based on alternatives like altmetrics are not free from this; Hammarfelt (2021) mentions that, despite being based on different criteria, these can also reflect the same biases, one of them being gender disparity in online visibility. The ethical consequence is unequivocal: if metric-based assessment systems do not incorporate a deliberate equity perspective, they will unintentionally preserve past injustices and thus, will not contribute to the establishment of a research culture that is more inclusive and representative.

4. Pathologies of Use: Gaming, Misconduct, and the Erosion of Integrity

Linking high-stakes decisions to certain indicators generates very strong motives for manipulations. The result is that there is a whole list of unethical behaviors that take place and that somehow affect the scientific record and distort the whole rationale behind the researchers' payment schemes.

4.1. *The Spectrum of Gaming and Manipulation*

Biagioli (2022) presents a nuanced analysis of the term "gaming metrics" and its comparison with blatant forms of scientific misconduct, such as data fabrication or falsification. Gaming is distinct from crime in that it uses the system's rules to boost one's academic profile, which often happens at the cost of the scientific communication's integrity. Among the practices that fall under this category is salami slicing, where a single research study is divided into several minimum publishable units, resulting in an artificial increase in publication count. Another method is strategic citing—acknowledging the works of colleagues, editors, or possible reviewers not due to academic relevance but to gain goodwill or to engage in reciprocal citation practices which make one's work more visible. Honorary authorship is also a widespread practice, where people who have a lot of influence are included in the author list to give it more weight or to get their help, even if they have not really contributed anything. Although these actions do not violate the rules technically, they still undermine the ethical standards of academic publishing. Götzsche (2021) points out that the constant demand for publication in high-impact journals has turned out to be the most important factor for scientific misconduct. Under such conditions, submitting shorter and less clear papers, overstating the results, or metric manipulation can be interpreted as tactics for surviving, thereby the pressing need for evaluation systems that value integrity over numerical performance is highlighted.

4.2. *The Journal Impact Factor and its Malignant Influence*

The Journal Impact Factor (JIF), though criticized a lot, still remains a strong and usually negative factor in the scientific research arena. Rushforth and de Rijcke (2020) used an ethnographic method to show how the JIF determines the course of biomedical research in the UK, affecting not only the scientific questions but also the ways of doing and the journals to be published in. This measure, which was primarily intended for journal assessment, is regularly abused to judge individual articles—an error-prone practice based on statistics that simply impairs proper valuation. Besides, the JIF turns the editorial conduct upside down by making the journals more concerned about highly cited review articles and thus rising the discouragement of putting out null results or replication studies. Such manipulation contributes significantly to a biased picture of scientific progress. The quest for the so-called high-JIF publications also makes the researchers' focus on trendy and low-risk topics more of an expectation than a choice, leaving behind the innovative or socially relevant ones. Ioannidis (2022) contends that such misuse has actually caused science to suffer, by promoting the replication crisis, silencing daring inquiries, and creating a highly competitive atmosphere. The ethical considerations are huge: the continuous use of the JIF feeds back a very narrow and distorted picture of the scientific value, thus, triggering a debate on the research quality measurement and reward methods that are used in today's academia.

4.3. *Over-Optimization and the Law of Diminishing Returns*

Fire and Guestrin (2021) highlight a troubling trend in academic publishing known as "over-optimization," where the relentless pursuit of metric maximization undermines the broader goals of scientific inquiry. This phenomenon exemplifies Goodhart's Law: when a measure becomes a target, it ceases to be a good measure. As researchers become increasingly skilled at gaming the system—strategically tailoring their work to boost citation counts, h-indexes, or journal impact factors—the effort required to achieve visibility and recognition escalates. Meanwhile, genuine scientific progress may stagnate, buried under a deluge of incremental publications that offer little substantive advancement. The peer-review system becomes overwhelmed, and researchers face mounting pressure and burnout. The ethical cost of this dynamic is profound: it represents a misallocation of intellectual energy and institutional resources, diverting attention away from bold, long-term, or socially transformative research. Instead of fostering innovation and discovery, the system incentivizes conformity and superficial productivity. Fire and Guestrin's analysis calls for a re-evaluation of how success is defined and measured in academia, urging stakeholders to prioritize meaningful contributions over metric performance and to design evaluation systems that support rather than distort the scientific enterprise.

5. New Frontiers and Emerging Ethical Challenges

The progress of academic communication along with data science still keeps on creating new measures and methods, where each one of them carries a distinct ethical consideration.

5.1. *Altmetrics: The Promise and Peril of Broader Impact*

Altmetrics, which come from social media, news sources, and many other digital places, might be the way to go when looking at the research impact through the societal lens that transcends academic borders. They might be the ones to ethically emphasize the work that profoundly impacts public discussions, policies, or even clinical practice, despite being scarcely cited. This wider acceptance at the DORA principle level that supports the different ways of scholarly impact evaluation.

Anyway, Haustein, Bowman, and Costas (2020) point out the difficulty of interpreting altmetrics as one of the drawbacks. Hammarfelt (2021) delineates the ethical dilemma of the interplay between visibility and validity mentioning that even though altmetrics have the capacity to increase the visibility of research, they might not always be the embodiment of its real worth. The norm of citations was that they would accumulate slowly and over a long period; they would be considered as part of the scholarly communication. But, in the case of altmetrics, they can be instantaneously ballooned thanks to bots, organized campaigns, or misleadingly attention-grabbing headlines. This, on one hand, goes beyond the limitation set by manipulation and creates a new bias, which now favors the research from institutions with powerful media or trendy topics. Consequently, the ethical use of altmetrics requires a critical and contextual approach. They are to be viewed as complementary indicators rather than sole measures of quality or impact. Without careful interpretation, altmetrics will continue to amplify the issue of lack of engagement and conceal the important contributions to science and society.

5.2. Algorithmic Evaluation and the Black Box

The integration of big data and machine learning into research assessment has introduced powerful algorithmic tools, but these innovations come with significant ethical concerns. Fire and Katz (2022) offer a data science perspective, emphasizing that algorithms trained on historical bibliometric data are prone to inheriting and amplifying existing biases. These include geographical, disciplinary, and gender-based disparities that have long shaped academic visibility and recognition. When such algorithms are deployed to evaluate job applications or funding proposals, they risk systematically disadvantaging entire groups of researchers, all under the guise of technological neutrality. Thelwall (2023) deepens this critique by examining the citation databases that feed these systems, revealing how selective coverage and structural biases in these sources further distort algorithmic outputs. The ethical imperative, therefore, is clear: research assessment must be governed by principles of transparency, auditability, and human oversight. Without these safeguards, algorithmic evaluation can entrench inequities and obscure the nuanced, context-rich dimensions of scholarly work. Responsible use of machine learning in academia demands not only technical rigor but also a commitment to fairness and inclusivity, ensuring that automation enhances rather than undermines the integrity of research evaluation.

5.3. The Ethics of Visualization and Communication

As scientific research grows in complexity, the visualization of bibliometric data has become a vital tool for interpreting and communicating scholarly trends. Techniques such as science maps, network graphs, and interactive dashboards help distill vast datasets into accessible formats. However, as Costas and van Eck (2022) emphasize, these visualizations carry significant ethical implications. Decisions about which data sources to include—such as relying solely on Web of Science—how to normalize the data, and how to design visual elements like color schemes or thresholds are far from neutral. These choices can subtly shape the viewer's perception, reinforcing dominant narratives of excellence while marginalizing less visible forms of research or institutions. For example, visual emphasis on highly cited work may obscure valuable contributions from underrepresented regions or disciplines. Ethical visualization demands reflexivity and transparency, ensuring that representations of science reflect its true diversity and complexity. It calls for a deliberate effort to avoid oversimplification and to resist the allure of aesthetic clarity at the expense of inclusivity. By acknowledging the interpretive power of visual design, researchers and evaluators can foster more equitable and accurate portrayals of the global scientific landscape.

6. Pathways to Reform: Towards a Culture of Responsible Assessment

Identifying the issues is the sole initial move. The implementation of systemic reform is the bigger difficulty. The literature suggests a multi-level strategy with the participation of researchers, institutions, funders and publishers.

6.1. Institutional Leadership and the Development of Local Frameworks

Institutions play a pivotal role in shaping the ethical landscape of research assessment, as they are responsible for critical decisions around hiring, promotion, and funding. In recent years, there has been a growing movement urging institutions to take proactive leadership in implementing responsible metrics policies. Gadd (2023) and Gadd and Fry (2023) offer practical guidance and case studies that illustrate how such policies can be developed and embedded within academic structures. This includes making a public commitment to reform by signing and adhering to the principles of DORA, revising promotion and tenure guidelines to de-emphasize journal-based metrics, and adopting a narrative-based approach that allows researchers to demonstrate a wide range of contributions. Equally important is the investment in training and capacity building, ensuring that administrators, hiring committees, and researchers are educated in the ethical use of metrics, as advocated by the Leiden Manifesto. The significance of institutional support is underscored by Berghaeuser, Prass, and Lindner (2025), whose pan-European survey reveals that researchers' ethical practices are strongly influenced by the clarity and consistency of guidance provided by their institutions. Robust support systems not only foster ethical behavior but also help cultivate a research culture grounded in fairness, transparency, and inclusivity.

6.2. The Centrality of Qualitative Judgement and Narrative CVs

The primary objective of reform in research evaluation is not to find a perfect measure, but to give expert judgment back its leading role. According to De Rijcke and Rushforth (2021), ethical evaluation has to evolve from a measuring process

to a judging one, underlining the subtle comprehension which only human evaluators can furnish. Tools like narrative CVs and contributor role taxonomies such as CRediT are facilitating this transition by allowing researchers to share their achievements, contributions to the field, and future plans in their own language. Such formats break away from numeral markers and provide a more elaborate and individualized account of the scholar's impact. One such case in point is the UK's *Résumé for Research and Innovation (R4RI)*, where applicants are invited to outline their work in terms of knowledge creation, mentoring, social impact and so on. Moher, Naudet, et al. (2021) are proponents of this method in hiring and promotion, stating its moral advantages. It is comprehensive, taking into account all types of academic work; contextual, being aware of differences in disciplines and career stages; and finally, it is supportive of various professional roads being seen as worthy of recognition. By using narrative and qualitative assessment, institutions may create an equitable and valuable system that acknowledges the complexity of research and the wide range of scholars' contributions to the community.

7. Conclusion

The exploration of ethics and bias in research metrics leads to a setting where a decision has to be made. The instruments that were made to facilitate the comprehension of the intricate enterprise, in most instances, have turned out to be the source of new and serious issues. They have demonstrated their vulnerability to geographical, disciplinary and gender biases; and they have produced the effects of gaming and misconduct along with their uncritical application, which has led to a loss of trust and distortion of scientific priorities. Nevertheless, this review also outlines a definite direction for the future. The answer does not involve a total rejection of all quantitative data like a Luddite but rather a mature, reflexive, and ethically-informed approach to its use. The foundational principles of DORA and the Leiden Manifesto are the basis for the blueprint. The reform movement is being reinforced, with a coalition of concerned scientists, forward-thinking institutions, and major funders supporting it. It is a movement towards evaluation that is:

- **Qualitative-led**, where metrics inform but do not replace expert judgment.
- **Holistic**, valuing the full range of research activities and impacts.
- **Context-aware**, sensitive to disciplinary norms, career stage, and institutional mission.
- **Equitable**, actively designed to identify and mitigate systemic biases.
- **Transparent**, open to scrutiny and dialogue.

Researchers like Adjovi (2025) and Berghaeuser et al. (2025) have provided empirical evidence that this matter should not only be seen as a technical issue but also as a cultural one; therefore, it will demand continuous education, institutional support, and a global discourse on the values we are willing to support in science. The main ethical requirement is to make sure that our evaluation methods promote a research culture that is characterized by strictness, creativity, and diversity, which will, in turn, depend on the public trust being the ultimate goal. If we strategically align our metrics with our mission, then we can start counting what really matters to us, instead of valuing en-masse what is easy to measure.

References

- Adjovi, I. S. M. (2025). *A worldwide itinerary of research ethics in science for a better social responsibility and justice: a bibliometric analysis and review*. *Frontiers in Research Metrics and Analytics*. [Link](#)
- Bar-Ilan, J. (2020). *The role of metrics in research assessment: Ethical considerations*. *Scientometrics*.
- Berghaeuser, H., Prass, M., & Lindner, R. (2025). *Researchers' practice and perception of research ethics and the role of institutional support: Insights from a pan-European survey*. *Journal of Academic Ethics*. [Link](#)
- Biagioli, M. (2022). *Gaming metrics: Misconduct and manipulation in scholarly publishing*. *Social Studies of Science*.
- Bornmann, L. (2021). *Normalization in citation metrics: Ethical implications and methodological challenges*. *Scientometrics*.
- Bouter, L. M. (2020). *Fostering responsible research practices: Metrics and ethics*. *Science and Engineering Ethics*.
- Costas, R., & van Eck, N. J. (2022). *Ethical visualization of science: Mapping responsibly*. *Quantitative Science Studies*.
- Cronin, B. (2020). *Citation ethics: The moral economy of referencing*. *Journal of Documentation*.
- De Rijcke, S., & Rushforth, A. (2021). *Ethics of research evaluation: Institutional perspectives*. *Research Evaluation*.
- Didegah, F., & Thelwall, M. (2021). *Gender differences in citation impact across disciplines: A large-scale analysis*. *Journal of Informetrics*.
- DORA (2020). *San Francisco Declaration on Research Assessment*. <https://sfdora.org>
- DORA (2025). *Global progress on responsible metrics: Annual report*. <https://sfdora.org>
- Fire, M., & Guestrin, C. (2021). *Over-optimization of academic publishing metrics: Observations and recommendations*. *EPJ Data Science*.
- Fire, M., & Katz, G. (2022). *Bias in algorithmic research evaluation: A data science perspective*. *EPJ Data Science*.
- Gadd, E. (2023). *Responsible metrics: A practical guide for institutions*. *UK Council for Research Integrity*.
- Gadd, E., & Fry, J. (2023). *Responsible metrics in practice: Institutional case studies*. *Research Evaluation*.

- Gagliardi, A. R., & Dobrow, M. J. (2023). *Ethical frameworks for evaluating implementation science*. Implementation Science.
- Giaume, L., Lamblin, A., Pinol, N., Gignoux-Froment, F., & Trousselard, M. (2025). *Evaluating cognitive bias in clinical ethics supports: A scoping review*. BMC Medical Ethics. [Link](#)
- Götzsche, P. C. (2021). *Scientific misconduct and the role of metrics*. Accountability in Research.
- Hammarfelt, B. (2021). *The ethics of altmetrics: Visibility vs. validity*. Quantitative Science Studies.
- Hammarfelt, B., & de Rijcke, S. (2021). *Accountability in research evaluation: Metrics and their moral dimensions*. Science and Engineering Ethics.
- Haustein, S., Bowman, T. D., & Costas, R. (2020). *Interpreting altmetrics: Viewing acts on social media through the lens of citation and social theories*. Quantitative Science Studies.
- Hicks, D. (2022). *Responsible research evaluation: Lessons from the Leiden Manifesto*. Nature Index.
- Hicks, D., & Wouters, P. (2020). *Ethical use of metrics in interdisciplinary research*. Scientometrics.
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2022). *Bibliometrics: The Leiden Manifesto for research metrics*. Nature.
- Ioannidis, J. P. A. (2022). *The misuse of metrics in science: A call for reform*. PLoS Medicine.
- Lane, J. (2020). *Metrics for science policy: Ethical dilemmas and design challenges*. Science and Public Policy.
- Larivière, V., & Sugimoto, C. R. (2022). *Global disparities in research metrics: Ethical implications*. Nature Human Behaviour.
- Larivière, V., Haustein, S., & Mongeon, P. (2020). *The oligopoly of academic publishers in the digital era*. PLoS ONE.
- MacRoberts, M. H., & MacRoberts, B. R. (2021). *Problems of citation analysis: Ethics and accuracy*. Journal of the Association for Information Science and Technology.
- Mingers, J., & Leydesdorff, L. (2020). *A review of the ethical dimensions of bibliometric indicators*. Journal of the Association for Information Science and Technology.
- Moher, D., & Goodman, S. N. (2022). *Ethical peer review and the influence of metrics*. PLoS Biology.
- Moher, D., & Naudet, F. (2021). *Ten principles for responsible research assessment*. PLoS Biology.
- Moher, D., Naudet, F., Cristea, I. A., Miedema, F., Ioannidis, J. P. A., & Goodman, S. N. (2021). *Assessing scientists for hiring, promotion, and tenure*. PLoS Biology.
- Mongeon, P., & Paul-Hus, A. (2022). *The ethics of database coverage: Implications for global research visibility*. Scientometrics.
- Neylon, C. (2022). *Ethics of research metrics: A framework for responsible use*. Research Evaluation.
- Paul-Hus, A., & Desrochers, N. (2021). *Gender bias in citation practices: A bibliometric perspective*. Scientometrics.
- Rafols, I. (2021). *Diversity and bias in research metrics: A systems view*. Research Policy.
- Rafols, I., & Yegros-Yegros, A. (2020). *Bias in journal rankings and its impact on interdisciplinary research*. Research Policy.
- Robinson-Garcia, N., & van Leeuwen, T. N. (2022). *The ethics of citation analysis: A review of practices and pitfalls*. Journal of Informetrics.
- Rushforth, A., & de Rijcke, S. (2020). *Accounting for impact? The Journal Impact Factor and the making of biomedical research in the UK*. Minerva.
- Sandstrom, U., & van den Besselaar, P. (2021). *Funding decisions and metric bias: An ethical analysis*. Research Evaluation.
- Sugimoto, C. R. (2023). *Equity in scholarly metrics: A feminist critique*. Journal of Informetrics.
- Sugimoto, C. R., & Larivière, V. (2021). *Equity in scholarly publishing: A global perspective*. Nature Human Behaviour.
- Tennant, J. P. (2023). *Open metrics and ethical transparency in scholarly communication*. F1000Research.
- Thelwall, M. (2023). *Algorithmic bias in citation databases: A critical review*. Journal of Informetrics.
- Whitley, R. (2020). *The social organization of science and the ethics of evaluation*. Minerva.
- Wilsdon, J. (2023). *Reforming research evaluation: Ethics, equity, and engagement*. UK Research Integrity Office.
- Wilsdon, J., Allen, L., Belfiore, E., et al. (2021). *The Metric Tide: Independent Review of the Role of Metrics in Research Assessment and Management*. UK Higher Education Funding Council.
- Wouters, P., & Costas, R. (2021). *Indicators of research quality: Beyond the impact factor*. Quantitative Science Studies.



© 2025 by the authors; licensee Growing Science, Canada. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).