International Science and Technology Evaluation Reform: A Ten-Year Review (Postprint)

Authors: Xu Fang, Li Xiaoxuan

Date: 2024-03-27T00:00:00+00:00

Abstract

May 2013, marked by the release of the San Francisco Declaration on Research Assessment (DORA), heralded the beginning of international reform in scientific and technological evaluation. Over the past decade, following extensive theoretical preparation and advocacy, this reform has progressed toward practical implementation, with over 350 institutions from more than 40 countries signing the Agreement on Reforming Research Assessment and launching pilot initiatives, yielding preliminary results. This article systematically summarizes and analyzes the decade-long international reform of scientific and technological evaluation from three aspects: reform objectives, measures, and case studies, and derives research conclusions and insights from a comparative perspective with China' s scientific and technological evaluation reform.

Full Text

Preamble

Citation: Xu Fang, Li Xiaoxuan. Review on reform of international research evaluation in past decade. *Bulletin of Chinese Academy of Sciences*, 2024, 39(1): 121-130, doi: 10.16418/j.issn.1000-3045.20231217002.

ChinaXiv: Review on Reform of International Research Evaluation in Past Decade
XU Fang, LI Xiaoxuan*

Over the past decade, international research evaluation reform has progressed from extensive theoretical preparation and advocacy to practical implementation. More than 350 institutions from over 40 countries have signed the Agreement on Reforming Research Assessment and launched pilot reforms, with initial results becoming evident. This article systematically summarizes and analyzes the ten-year international research evaluation reform from three perspectives:

reform objectives, measures, and case studies, and draws research conclusions and insights through comparison with China' s research evaluation reform.

Keywords: international, research evaluation, reform, San Francisco Declaration on Research Assessment (DORA)
DOI: 10.16418/j.issn.1000-3045.20231217002
CSTR: 32128.14.CASbulletin.20231217002

---

*XU Fang is Deputy Director of Management Innovation and Research Evaluation Center of Chinese Academy of Sciences (CAS), Associate Professor of Institutes of Science and Development, CAS, and Member of the International Network of Research Management Societies (INORMS) Research Evaluation Working Group. Her main research interests include S&T management and evaluation, S&T policy, and she is primarily responsible for the evaluation of CAS institutes. E-mail: xufang@casisd.cn

LI Xiaoxuan is Director of Management Innovation and Research Evaluation Center of Chinese Academy of Sciences (CAS), Professor of Institutes of Science and Development, CAS, and Executive Member of the Declaration on Research Assessment (DORA). His research focuses on scientific research management, including research evaluation, human resource management, and research funding management. E-mail: xiaoxuan@casisd.cn

---

1. What Problems Does International Research Evaluation Reform Aim to Solve?

Generally speaking, traditional scientific powers in Europe and North America possess strong scientific cultural foundations and have traditionally upheld sound scientific values and pursuits. However, with the development of science and technology and their rising strategic importance, these countries' research evaluation systems face new problems and challenges that require timely responses. Broadly speaking, international research evaluation reform aims to address three major issues.

First, avoiding the misuse of bibliometric methods in research evaluation. After Garfield [3] proposed citation analysis for tracking scientific progress, the Institute for Scientific Information (ISI) began selecting journals through citation analysis in 1963, forming the Science Citation Index (SCI) database that provided the foundation for applying bibliometric methods in research evaluation. While bibliometric methods have provided evidence-based support for evaluation, they have also fueled the gradual rise of "evaluating papers by journals and evaluating people by papers" —the notion that where one publishes matters more than what one publishes. This undoubtedly exerts negative effects on the quality, integrity, and diversity of research output. How to prevent the misuse

of bibliometric methods has become a significant challenge for the international scientific community.

Second, emphasizing evaluation of science' s impact on economy and society. As science and technology have gained prominence in national economic and social development and national security, global competition in science and technology has intensified. Countries worldwide have increased S&T investment while paying greater attention to the efficiency and effectiveness of such investment in driving domestic innovation development. The linear model of science investment that "only asks about cultivation, not about harvest" has been questioned, and impact evaluation has gradually become central to research evaluation. However, this shift presents two challenges: the scientific community struggles to reach consensus, with many researchers rejecting impact evaluation as vaguely defined and prone to self-aggrandizement, potentially encouraging research misconduct and damaging academic culture; and accurately assessing impact is extremely difficult, making it hard to identify scientific indicators, data sources, and evaluation methods. Both issues remain hot topics in international research evaluation discussions.

Third, adapting to new paradigms such as open science and AI-driven scientific research. Open science, based on data sharing, has gained momentum in Europe and America in recent years and is gradually influencing the world. The open science movement demands reform of research evaluation systems to enhance openness and transparency. However, transforming research evaluation from focusing on individual creative activities to recognizing collective research activities characterized by data sharing and large-scale collaboration requires concerted efforts from all stakeholders. For instance, UNESCO has issued recommendations on open science [5], including an "Open Science Toolkit" to help members review and reform research career assessment standards. The rapid development of artificial intelligence will also profoundly impact research evaluation, with "AI for Science" (AI4S) emerging as a new scientific paradigm [6]. Countries are striving to seize the commanding heights of this new paradigm and need to incentivize and guide it through research evaluation. Meanwhile, while AI4S promotes scientific development and reduces scientists' workload, it may also reinforce risks and biases associated with data prediction technologies, posing new challenges for reforming research evaluation. However, although frequently mentioned, this has not yet become a focus of the past decade' s international research evaluation reform.

Regarding these three aspects, the urgent problems or core objectives of this international research evaluation reform are the first two: the misuse of bibliometric methods and impact evaluation. These align with China' s situation. The first problem—the misuse of bibliometric methods—is particularly severe in China. Compared with traditional scientific powers, China' s peer review system is less robust due to a weaker scientific culture and excessive influence of personal connections, leading to greater reliance on quantitative indicators such as journal impact factors, citation counts, and paper numbers. The second

problem—how to promote impact evaluation—is similar to China' s current push for the "five-dimensional value" evaluation of S&T achievements [28]. However, China' s efforts to break the "Four Onlys" —focusing on awards, academic credentials, professional titles, and talent "hats" —are more characteristic of China' s specific context. For many traditional scientific powers, these fundamental evaluation issues are not as pronounced.

2. What Measures Have Been Taken in International Research Evaluation Reform?

2.1 Major Initiatives in International Research Evaluation Reform

In terms of approach, international research evaluation reform has been led by the scientific community, primarily through bottom-up methods. The reform' s launch was marked by the release of the San Francisco Declaration on Research Assessment (DORA) in May 2013. The declaration' s draft was developed by scholars and editors during the 2012 American Society for Cell Biology (ASCB) annual meeting in San Francisco to address the misuse of journal impact factors in research evaluation. Following DORA' s release, numerous international academic organizations, conferences, and universities/research institutions joined the cause, establishing new organizations such as the DORA Research Evaluation Alliance and the INORMS Research Evaluation Working Group to promote reform. In May 2023, many countries worldwide, including China, held commemorations of DORA' s 10th anniversary.

Over the past decade, the international scientific community has undertaken extensive and diverse efforts to promote research evaluation reform, including issuing declarations, initiatives, and statements; organizing academic conferences, thematic discussions, and research projects; and producing research reports, evaluation frameworks, best practice cases, and pilot agreements. This article identifies 14 particularly significant measures (Table 1).

2.2 Main Effects of International Research Evaluation Reform

First, a global consensus on research evaluation reform has formed. As of January 4, 2024, 3,078 organizations and 21,339 individuals have signed DORA, including 15 institutions from China. In 2022, the Coalition for Advancing Research Assessment (CoARA) was formally established and released the Agreement on Reforming Research Assessment, which has been signed by over 350 organizations from more than 40 countries. Research evaluation reform is increasingly gaining global consensus.

Second, through the concerted efforts of the scientific community, the "map" of research evaluation reform has gradually become clearer. For example, DORA proposed eliminating "evaluating papers by journals" ; the Leiden Manifesto further called for correcting "quantitative evaluation" ; The Metric Tide report further clarified the role and norms of quantitative evaluation; and the SCOPE

framework defined the process of responsible evaluation. Different academic organizations have addressed different aspects of research evaluation reform, piecing together a relatively complete "map." This reform "map" has been labeled as "responsible research assessment" and is gradually becoming common terminology in the scientific community.

Third, research evaluation reform is moving from conceptualization to practice. Currently, the more than 3,000 organizations that have signed DORA are implementing or have implemented requirements to avoid "evaluating papers by journals." The over 300 organizations (including funding agencies, universities, and research institutions) that have signed the Agreement on Reforming Research Assessment are conducting pilot reforms and regularly organizing various forms of experience-sharing activities.

Fourth, several fundamental judgments about research evaluation have emerged, including understanding its positive and negative roles, the relationship between quantitative and qualitative evaluation, prerequisites for launching evaluations, and techniques for improving evaluation data quality. These rational insights offer important lessons for China.

3. International Practice Cases of the "Three Evaluations" Reform

As mentioned above, international research evaluation reform is transitioning from theory to practice. The following case analyses are presented. Given that China's current research evaluation reform originates from national "Three Evaluations" reform documents, this section selects cases covering talent evaluation, project review, and institutional assessment.

3.1 Talent Evaluation Reform at Ghent University

Ghent University in Belgium was among the first to recognize the systematic damage to research culture caused by quantitative evaluation based on bibliometric methods, believing that such evaluation fostered a culture where "where one publishes matters more than what one publishes." After DORA' s release in 2013, Ghent University signed the declaration. Subsequently, it also signed the Agreement on Reforming Research Assessment. The university then undertook reforms of faculty promotion and other talent evaluation systems to preserve its advocated culture of diversity, address growing faculty dissatisfaction with quantitative evaluation, and create a career framework that emphasizes the shared value of pursuing excellent research—challenging, high-quality, and motivating.

Through collaborative efforts between university management and faculty, Ghent University released its Vision Statement for Evaluating Research in November 2016 [21], proposing eight principles that research evaluation must follow. In 2017, it further published guidelines for using quantitative indicators in research evaluation [22]. Based on these two policies, Ghent University

established a completely new faculty evaluation and promotion model in 2018, returning "responsibility" and academic freedom to professorial staff. Under the new system, faculty evaluation at Ghent University no longer focuses solely on research output but adopts a more qualitative, comprehensive, and human-centered perspective. Evaluations follow a five-year cycle, including initial evidence-based assessment, mid-term feedback interviews, and final interview-based evaluation. The evaluation content includes narrative presentations of the most important achievements in research, teaching, societal engagement, and management/leadership (rather than using measurable quantitative criteria), along with plans for the next five years.

3.2 Project Review Reform at the U.S. National Institutes of Health (NIH)

The U.S. National Institutes of Health (NIH) is the primary medical research and funding agency in the United States, annually funding numerous new projects to advance relevant fields. After signing DORA, NIH undertook project review reforms to eliminate existing quantitative issues and biases in peer review while adapting to open science development. The reform 主要包括 three aspects:

First, modifying review rules [23]. The new rules require that evaluations of researchers and research environments must be considered within the context of the research project rather than being scored separately as before. For both personnel and institutions, the review standard is no longer "the stronger the better" but "adequate to accomplish the task" ; if reviewers believe personnel or institutional capacity is insufficient, they must provide specific justifications. This "adequacy" principle attempts to address prestige bias by focusing more attention on the research topic itself rather than institutional reputation.

Second, revising the biographical sketch or "biosketch" format used in project applications [24]. A brief section was added to the biosketch where applicants describe their most significant scientific achievements, shifting reviewers' attention away from the journals in which previous papers were published.

Third, introducing new data management and sharing policies. Since January 2023, most of the 300,000 researchers and 2,500 institutions receiving NIH funding annually have been required to articulate data management and sharing (DMS) plans in their grant applications [25]. DMS plans should include details about software or tools needed to analyze data, when and where raw data will be made available, and any special considerations for accessing or distributing the data, with justifications provided for any restrictions or exceptions to data sharing, thereby promoting open science.

3.3 University Evaluation Reform in the United Kingdom

In 2014, the UK substantially reformed its previous Research Assessment Exercise (RAE) system, creating the new Research Excellence Framework (REF).

Compared to RAE, REF' s most significant innovation was introducing bibliometric indicators to provide reference for peer review and exploring impact evaluation methods [26] to demonstrate UK universities' real societal impact, emphasizing the tangible benefits that scientific research brings to the real world. Since impact evaluation is particularly challenging, the UK conducted specialized research to develop impact indicators for different disciplinary achievements.

In 2015, commissioned by the UK Higher Education Funding Council, Professor James Wilsdon' s research group conducted an independent assessment of the role of quantitative indicators in REF. The group launched the Responsible Metrics Forum to focus on and discuss how to properly use quantitative indicators in research evaluation. The research group ultimately published The Metric Tide report, offering positive assessments of quantitative indicator usage and proposing improvement recommendations.

In 2022, Research England—the lead agency for REF—signed the Agreement on Reforming Research Assessment, committing to further reform REF to reestablish a responsible, inclusive, and diverse research culture across the entire university research ecosystem. Consequently, the UK launched the Future Research Assessment Programme [27] to conduct more in-depth research on future reform plans for REF 2021. According to the top-level design plan released in June 2023, REF 2028 will shift its policy focus from "research performance incentives" to "research culture building," comprehensively reshaping the three evaluation dimensions of research environment, research output, and research impact to enhance universities' emphasis on building a healthy research culture.

3.4 Main Experiences from International "Three Evaluations" Reform Practice

First, integrating theory and practice. Pilot institutions fully utilize the theoretical and methodological frameworks developed by the international scientific community during the decade of research evaluation reform, maintaining close integration between theory and practice.

Second, maintaining communication and sharing among pilot institutions. Pilot institutions generally sign both DORA and the Agreement on Reforming Research Assessment, positioning themselves within a collective of reformers and maintaining interactive communication with research evaluation scholars and other pilot institutions.

Third, continuously innovating and improving evaluation methods rather than expecting immediate perfection. For example, REF began exploring research impact evaluation in 2008, investing substantial human and time resources to develop the concept, evaluation criteria, methods, and expert handbooks for impact assessment before it was applied in actual evaluation in 2014—and research and improvement continue today.

Fourth, pilot institutions demonstrate full reform autonomy. Institutions have

launched reforms in response to international scientific community initiatives based on their own philosophical 认同 and actual needs, entirely autonomously rather than through government administrative requirements. While the evaluation results inevitably transmit competitive pressure through universities to grassroots academic organizations, particularly affecting individual faculty research behavior, the introduction of quantitative indicators has exacerbated this impact.

4. Conclusions and Implications

4.1 Conclusions

Much can be learned from the decade of international research evaluation reform. From the perspective of comparison with China's reform, this article draws three main conclusions:

First, international research evaluation reform shares similar goals with China' s reform. The core objectives of this international reform are twofold: (1) breaking "evaluating papers by journals," which aligns with China' s orientation of breaking the "Four Onlys" , particularly breaking the excessive reliance on papers; and (2) establishing "impact" evaluation for economic and social contributions, which aligns with China' s emphasis on the five-dimensional value of S&T achievements [28]. However, China' s efforts to break the "Four Onlys" — focusing on awards, academic credentials, professional titles, and talent "hats" —are more characteristic of China' s specific context. For traditional scientific powers, these fundamental evaluation issues are not as pronounced.

Second, international research evaluation reform follows a significantly different path from China' s reform. International reform is primarily led by the scientific community through bottom-up approaches, using declarations, initiatives, evaluation frameworks, signed commitments, and case studies to drive change, with minimal direct government intervention. In contrast, China employs a more top-down approach, with the government playing a leading role in driving reform through policy documents and requirements, while the scientific community' s role remains limited.

Third, the decade of international research evaluation reform offers valuable lessons. Unlike China' s top-down reform with strong execution power, international reform places greater emphasis on understanding the fundamental principles of research evaluation. Consequently, international reform has achieved relatively systematic consensus worth learning from. Of course, many consensuses reached through international reform align with concepts China has upheld in practice, with some showing remarkable similarities. International consensus includes: DORA' s 倡议 to focus on papers themselves rather than journals and its recommendations for different stakeholders; the Leiden Manifesto' s 倡议 on proper use of quantitative evaluation; the SCOPE framework proposed by the INORMS Research Evaluation Working Group; and the impact evaluation methods developed through the UK' s REF. Several core perspectives

deserve special mention: research evaluation is necessary for management but is a double-edged sword that should not be launched without sufficient reason and preparation; research evaluation cannot rely on simple quantitative methods and must employ peer review, though quantitative methods can effectively enhance peer review quality when used appropriately; quantitative methods vary in quality and must be distinguished and selected carefully in practice—good methods like the Category Normalized Citation Impact (CNCI) indicator [29] should be used, while poor methods like journal impact factors should be avoided; attention must be paid to data quality and source reliability for quantitative evaluation, as attractive indicators are useless without quality and reliability; and using unique identifiers (such as ORCID) to link research outputs and activities with researchers has become increasingly common practice in the international scientific community, facilitating data access while ensuring quality.

4.2 Implications

The conclusions drawn from international comparison offer many implications for China' s research evaluation reform. This article focuses on four key implications:

First, break "excessive reliance" categorically and step-by-step. Reform must clarify responsible entities and sequencing. Among China's current "Four Onlys," different types of "excessive reliance" require different approaches: (1) Severity varies—"excessive reliance on papers,""awards,"and "talent hats"are more serious and require focused attention, while "academic credentials" and "professional titles" are also problematic but affect smaller groups; (2) Responsible entities differ—"excessive reliance on papers"primarily concerns the scientific community, aligning with international reform, while "awards" and "talent hats" primarily concern the government, an issue not present in international reform. These two responsible entities interact, but there is a logical sequence. The problems of "awards" and "talent hats" stem from their excessive numbers and require the government to reduce them, providing foundational institutional guarantees for the scientific community to break "excessive reliance on papers" and focus on producing original achievements. Regarding "excessive reliance on papers," China should draw on international experience to better mobilize the scientific community' s bottom-up reform initiative, creating a situation where pioneers dare to be the first to try new approaches.

Second, be cautious about launching evaluations. Evaluation is a double-edged sword. International research evaluation reform advocates against launching research evaluation lightly, which resonates with the pace of China' s reform. In fact, before breaking the "Four Onlys,"China first undertook "reduction"reforms, requiring institutions to eliminate excessive, frequent, and overlapping evaluations. However, after the reduction momentum faded, the impulse to evaluate has resurfaced. Additionally, the requirement for "comprehensive implementation of budget performance management"has not yet been well coordinated with

existing research evaluation, leading to increased evaluation frequency. This necessitates establishing mechanisms to evaluate the evaluations themselves to constrain evaluation impulses and standardize evaluation systems and methods, rather than merely conducting reduction campaigns as temporary movements.

Third, make good use of quantitative evaluation. As an auxiliary method to peer review, quantitative evaluation has received focused attention in this international reform, generating considerable consensus. Given China' s past overreliance on quantitative evaluation, some argue for completely abandoning it and returning to pure peer review. Considering China' s national conditions, this is not advisable. The author previously proposed the BRIDGE theory combining quantitative and qualitative evaluation [30], advocating for form-based methods to make tacit knowledge from data and evidence explicit, thereby supporting and constraining peer review. This approach would both maximize the utility of China' s existing quantitative evaluation exploration and potentially achieve breakthroughs in combining quantitative and qualitative methods, contributing Chinese innovations to international research evaluation reform.

Fourth, actively integrate into international research evaluation reform. Currently, relatively few Chinese research institutions, universities, and individual scientists have signed DORA, which is disproportionate to China' s vast scientific community. Moreover, no Chinese research institutions or universities have joined the international Agreement on Reforming Research Assessment. This situation relates to China' s status as a latecomer nation still gradually integrating into the international scientific community, the government-led nature of China' s research evaluation reform, and even the impact of the recent pandemic. As part of the international scientific community, China should more actively integrate into international research evaluation reform. Through mutual learning and promotion with the international scientific community, China can better stimulate its scientific community' s initiative in research evaluation reform while enhancing mutual understanding and trust, thereby strengthening bonds for comprehensive international S&T cooperation.

References

1. Xu F, Jin X X, Lyu W J, et al. International science and technology evaluation reform' s practice and enlightenment—A exploration represented by DORA. *Journal of South China University of Technology (Social Science Edition)*, 2022, 24(3): 94-100. (in Chinese)

2. Xu F, Li X X. Review on reform of research evaluation in past decade. *Bulletin of Chinese Academy of Sciences*, 2022, 37(5): 603-612. (in Chinese)

3. Garfield E. Citation indexes for science: A new dimension in documentation through association of ideas. *Science*, 1955, 122: 108-111.

4. Sheng X P, Bi C C, Tang J J. A review of open science research at home

and abroad. *Documentation, Information & Knowledge*, 2022, 39(4): 101-113. (in Chinese)

5. United Nations Educational, Scientific and Cultural Organization. *UNESCO Recommendation on Open Science*. (2021-11-09) [2023-12-26]. https://unesdoc.unesco.org/ark:/48223/pf000037994.
6. Wang F Y, Miao Q H. Novel paradigm for AI-driven scientific research: From AI4S to intelligent science. *Bulletin of Chinese Academy of Sciences*, 2023, 38(4): 536-540. (in Chinese)
7. Hicks D, Wouters P, Waltman L, et al. Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 2015, 520: 429-431.
8. Wilsdon J, Allen L, Belfiore E, et al. *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. (2015-07-06) [2023-12-26]. https://www.ukri.org/wp-content/uploads/2021/12/RE-151221-TheMetricTideFullReport2015.pdf.
9. The HuMetricsHSS Team. Humane Metrics Initiative. [2023-12-26]. https://humetricshss.org/humetricshss-principles/.
10. Wilsdon J, Bar-Ilan J, Frodeman R, et al. Next-generation metrics: responsible metrics and evaluation for open science. (2017-03-20) [2023-12-26]. https://www.researchgate.net/publication/315385216_{Next}-generation_{{{metrics}}*{{responsible}}*}{{{metrics}}*{{and}}*}{{{evaluation}}*{{for}}*}{{{open}}_{{s
11. Moher D, Naudet F, Cristea I A, et al. Assessing scientists for hiring, promotion, and tenure. *PLoS Biology*, 2018, 16(3): e2004089.
12. INORMS-RWG. *SCOPE Framework for Research Evaluation*. [2024-01-04]. https://inorms.net/scope-framework-for-research-evaluation/.
13. European University Association. *EUA Roadmap on Research Assessment in the Transition to Open Science*. (2018-06-18) [2024-01-04]. https://eua.eu/downloads/publications/eua-roadmap-on-research-assessment-in-the-transition-to-open-science.pdf.
14. Dominik M, Badre A, Almas T A, et al. Publishing models, assessment, and open science. Report and outcomes from a workshop held by the Global Young Academy. (2018-10-29) [2023-12-26]. https://globalyoungacademy.net/wp-content/uploads/2018/10/APOS-Report-29.10.2018.pdf.
15. Global Research Council. *Statement of Principles on Peer/Merit Review*. [2023-12-26]. https://globalresearchcouncil.org/fileadmin//documents/GRC_{Publications}/French/GF
16. Moher D, Bouter L, Kleinert S, et al. The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLoS Biology*, 2020, 18(7): e3000737.

17. Science Europe. *Science Europe Position Statement on Research Assessment Processes*. (2020-02-04) [2023-12-26]. https://www.scienceeurope.org/our-resources/science-europe-study-on-research-assessment-practices/.

18. Global Research Council. *Responsible Research Assessment Conference Summary Report*—Global Research Council (GRC) Conference Report 2021. (2020-11-23) [2023-12-26]. https://globalresearchcouncil.org/fileadmin/documents/GRC_{Publica

19. Research and Innovation UK. *Reforming Research Assessment Agreement is now final*. (2022-07-20) [2023-12-26]. https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/reforming-research-assessment-agreement-now-final-2022-07-20_{en}.

20. Curry S, Rijcke S D, Hatch A, et al. The changing role of funders in responsible research assessment: Progress, obstacles and the way ahead. (2020-11-18) [2023-12-26]. https://rori.figshare.com/ndownloader/files/25518674.

21. Ghent University. *Vision Statement for Evaluating Research at Ghent University*. (2016-11-08) [2023-12-26]. https://www.ugent.be/en/research/research-strategy/evaluation/research-evaluation-principles.pdf.

22. Ghent University. *Tips for evaluating research and researchers*. (2023-10-9) [2024-01-04]. https://onderzoektips.ugent.be/en/tips/00001832/.

23. Kozlov M. NIH plans grant-review overhaul to reduce bias. *Nature*, 2022, 612: 602-603.

24. National Institutes of Health. *Biographical Sketch*. [2023-12-26]. https://grants.nih.gov/grants/forms/biosketch.html.

25. Kozlov M. NIH issues a seismic mandate: Share data publicly. *Nature*, 2022, 602: 558-559.

26. Xu F, Liu W B, Li X X. Research impact evaluation within UK REF and its implications. *Science of Science and Management of S & T*, 2014, 35(7): 9-15. (in Chinese)

27. Research and Innovation UK. *Future Research Assessment Programme*. (2022-12-09) [2023-12-26]. https://www.ukri.org/who-we-are/research-england/research-excellence/future-research-assessment-programme-frap/.

28. The State Council of China. *Guiding Opinions on Improving the Evaluation Mechanism for Scientific and Technological Achievements*. (2021-08-02) [2023-12-26]. https://www.gov.cn/zhengce/zhengceku/2021-08/02/content_{5628987}.htm?eqid=efd6cb8a00015759000000002646f098b.

29. Hao R Y. Study on correlations among three standardized evaluation indices: RCR, CNCI and JNCI. *Journal of Modern Information*, 2023, 43(12): 133-142. (in Chinese)

30. Xu F, Li X X. Between “Breaking”and “Building”: The BRIDGE Theory of research evaluation. *Bulletin of Chinese Academy of Sciences*, 2022, 37(8): 1099-1107. (in Chinese)

Note: China’ s “Four Onlys” (唯论文、唯奖项、唯学历、唯职称) refers to the problematic practice of evaluating researchers based solely on papers, awards, academic credentials, and professional titles. This was later expanded to “Five Onlys” with the addition of “talent hats” (唯帽子).

Note: Figure translations are in progress. See original paper for figures.