RoRI RESEARCH ON RESEARCH INSTITUTE

RoRI Working Paper No.15

A new typology of national research assessment systems: continuity and change in 13 countries

Alexander Rushforth, Gunnar Sivertsen, Ana Arango, Adriana Bin, Catriona Firth, Claire Fraser, Nino Gogadze, Natalia Gras, Lee Harris, Jon Holm, Peter Kolarz, Moumita Koley, Jorge Maldonado, Marie-Helene Nienaltowski, Laura Rovelli, Sergio Salles-Filho, Scipione Sarlo, Nerina Sarthou, Arne Sjostedt, Federico Vasen, Nicole Ward-Boot, James Wilsdon, Marta Wroblewska, Fang Xu, Lin Zhang

May 2025

About the authors

Alexander Rushforth is a senior research fellow at RoRI and Assistant Professor at the Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands. a.d.rushforth@cwts.leidenuniv.nl; ORCID 0000-0003-3352-943X; https://www.cwts.nl/people/AlexRushforth

Gunnar Sivertsen is Emeritus Professor at the Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway and a visiting senior research fellow at RoRI. gunnar.sivertsen@nifu.no; ORCID 0000-0003-1020-3189; https://www.nifu.no/ansatte/gunnar-sivertsen/

Ana María Arango-Hoyos is based at CoLaV, Universidad de Antioquia UdeA, Medellín, Colombia. annyarango@gmail.com

Adriana Bin is Assistant Professor at the School of Applied Sciences, State University of Campinas, Brazil. adribin@unicamp.br

Catriona Firth is Associate Director, Research Environment at Research England, UKRI. catriona.firth@re.ukri.org; ORCID 0000-0001-9927-0032.

Claire Fraser is a Senior Policy Advisor at Research England, UKRI. claire.fraser@re.ukri.org; ORCID 0009-0004-9984-1121.

Nino Gogadze is a research fellow at RoRI and a PhD candidate at the Centre for Science and Technology Studies (CWTS), Leiden University, the Netherlands. n.gogadze@cwts.leidenuniv.nl; ORCID 0009-0006-4498-9481; https://www.universiteitleiden.nl/en/staffmembers/nino-gogadze#tab-1.

Natalia Gras is Associate Professor with the University Research Council at the University of the Republic, Uruguay; nataliagras@gmail.com; ORCID 0000-0002-2653-8314.

Lee Harris is Director, Evaluation and Impact Delivery at the Australian Research Council (ARC) lee.harris@arc.gov.au

Jon Holm is a special adviser at the Research Council of Norway, jon.holm@rcn.no, ORCID 0000-0002-5345-2639.

Peter Kolarz is Head of Programmes at RoRI and an associate staff member in the Department of Science, Technology, Engineering and Public Policy (STEaPP) at University College London (UCL). ORCID 0000-0003-4443-580X; https://researchonresearch.org/team/peter-kolarz/

Moumita Koley is a research fellow at RoRI and senior research analyst at DST-CPR, Indian Institute of Science, Bangalore, India. moumitakoley@iisc.ac.in; ORCID 0000-0003-2394-0663; https://dstcpriisc.org/moumita-koley/

Jorge Maldonado Soto is a postgraduate student at Universidad Alberto Hurtado, Chile. jmaldona89@gmail.com; ORCID 0009-0007-9650-7605,

Marie-Helene Nienaltowski is a Senior Policy Advisor at Research England, UKRI, United Kingdom. marie-helene.nienaltowski@re.ukri.org; ORCID 0009-0001-4028-1770

Laura Rovelli is a research fellow at RoRI; former coordinator of CLACSO-FOLEC; and a researcher at National University of La Plata/CONICET, Argentina. laurarovelli@gmail.com; ORCID 0000-0001-7059-149X

Sergio Salles-Filho is professor at the Department of Science and Technology Policy, State University of Campinas, Brazil, sallesfi@unicamp.br; ORCID 0000-0002-8727-450; https://portal.ige.unicamp.br/en/faculty/sergio-luiz-monteiro-salles-filho

Scipione Sarlo is a Senior Evaluation Officer at ANVUR, Rome, Italy. scipione.sarlo@anvur.it; ORCID 0000-0002-2448-5220; https://www.anvur.it/it/persone/sarlo-scipione.

Nerina Sarthou is a researcher at the National University of Central Buenos Aires (UNICEN/CONICET), Argentina, nsarthou@fch.unicen.edu.ar; ORCID 0000-0003-2671-4725.

Arne Sjostedt is an Assistant Director for Research Evaluation and Impact Policy at the Australian Research Council. arne.sjostedt@arc.gov.au

Federico Vasen is researcher at the University of Buenos Aires/CONICET, Argentina, and a member of the Scholarly Communication Research Group at AMU, Poland. fvasen@rec.uba.ar; ORCID 0000-0001-9132-6206.

Nicole Ward-Boot is Monitoring, Evaluation and Learning Coordinator in the Strategy Department of Dutch National Research Council (NWO), n.ward-boot@nwo.nl https://www.linkedin.com/in/nicole-ward-boot-a8b4657/-

James Wilsdon is Executive Director of RoRI and Professor of Research Policy in the Department of STEaPP at University College London (UCL). j.wilsdon@ucl.ac.uk; ORCID 0000-0002-5395-5949; https://profiles.ucl.ac.uk/93009-james-wilsdon/about

Marta Natalia Wróblewska is Assistant Professor at SWPS University, Warsaw, Poland. mwroblewska2@swps.edu.pl; ORCID 0000-0001-8575-5215; https://martawroblewska.pl/

Fang Xu is a Professor in Institutes of Science and Development, Chinese Academy of Sciences (CAS), and deputy director of Research Evaluation Center of CAS. xufang@casisd.cn; ORCID 0000-0002-0622-6663; http://www.casisd.cn/sourcedb/zw/zjrc/201708/t20170815_4847083.html.

Lin Zhang is Full Professor, School of Information Management, Wuhan University, China. Zhanglin_1117@126.com; ORCID 0000-0003-0526-9677; https://sim.whu.edu.cn/info/1204/86172.htm

For citation: Rushforth, A., Sivertsen, G, Bin, A., Firth, C., Fraser, C., Gogadze, N., Gras, N., Harris, L., Holm, J., Kolarz, P., Koley, M., Maldonando Soto, J., Nienaltowski, H-M., Rovelli, L., Salles-Filho, S., Sarlo, S., Sarthou, N., Sjostedt, A., Vasen, F., Ward-Boot, N., Wilsdon, J., Wróblewska, M., Xu, F., Zhang, L. (2025) **A new typology of national research assessment systems: continuity and change in 13 countries.** RoRI Working Paper No. 15, May 2025. **DOI 10.6084/m9.figshare.29041787**

Acknowledgements

This paper forms part of AGORRA: A Global Observatory of Responsible Research Assessment, a collaboration led by the Research on Research Institute (RoRI).

RoRI's second phase (2023–2027) is funded by an international consortium of partners, including: Australian Research Council (ARC); Canadian Institutes of Health Research (CIHR); Digital Science; Dutch Research Council (NWO); FWF-Austrian Science Fund; Gordon and Betty Moore Foundation *[Grant number GBMF12312; DOI 10.37807/GBMF12312];* King Baudouin Foundation; La Caixa Foundation; Leiden University; Luxembourg National Research Fund (FNR); Michael Smith Health Research BC; National Research Foundation of South Africa; Novo Nordisk Foundation *[Grant number NNF23SA0083996]*; Research England (part of UK Research and Innovation); Social Sciences and Humanities Research Council of Canada (SSHRC); Swiss National Science Foundation (SNSF); University College London (UCL); Volkswagen Foundation; and Wellcome Trust *[Grant number 228086/Z/23/Z]*.

The Research Council of Norway has funded the research of Gunnar Sivertsen in this project through grant number 256223 R-Quest.

Sincere thanks to all AGORRA partners for their engagement and support:

- Research England/UKRI
- Australian Research Council
- ANVUR (Agenzia Nazionale Di Valutazione Del Sistema Universitario E Della Ricerca)
- Evaluation Research Centre, Chinese Academy of Sciences (CAS), Beijing
- Dutch Research Council (NWO)
- The National Research Foundation of South Africa (NRF-SA)
- Canadian Institutes of Health Research (CIHR)
- Social Sciences and Humanities Research Council (SSHRC)
- Swedish Research Council (SRC)
- Research Council of Norway (RCN)
- Volkswagen Foundation

We would also like to record our gratitude to members of the AGORRA working group for advice and guidance at every stage. Responsibility for the content of RoRI outputs lies with the authors and RoRI CIC. Any views expressed do not necessarily reflect those of our partners. RoRI is committed to open research as an important enabler of our mission, as set out in our Open Research Policy. Any errors or omissions remain our own.

RoRI RESEARCH ON RESEARCH INSTITUTE

researchonresearch.org/

Contents

Abstract

How are national systems for assessing publicly funded research evolving? What purposes do they serve and how are they designed to fulfil these? This working paper surveys the landscape of **national research assessment and funding systems across thirteen countries from 2010 to 2024,** and makes three contributions to our understanding of these systems.

First, we advance a **new typology** to categorize and compare important characteristics of these systems, providing insights into their similarities and differences, and a basis for mutual learning.

Second, we **identify and compare important shifts** over time across the thirteen systems through the framework of t**hree dynamic and interacting research performance paradigms.** These point to a gradual shift away from narrow conceptions of research 'excellence' towards more holistic criteria of value, qualities and impacts across several systems – though not all.

Finally, we consider **potential trajectories over the next decade**: including how a variety of assessment systems might respond to and incorporate **responsible research assessment (RRA) movements for reform.**

This paper is a **unique collaboration** between researchers and colleagues working in the funding or evaluation agencies of the thirteen countries in question. By mapping these research assessment systems and identifying dynamics of change, the paper offers **insights for policymakers, research funders and institutional leaders** looking to navigate this terrain at a time of heightened yet shifting expectations.

1. Introduction

This paper develops a novel typology for comparing national research assessment and funding systems worldwide, then analyses significant shifts over the past fifteen years.

The typology reflects patterns observed across national systems from thirteen countries. It forms part of **AGORRA: A Global Observatory of Responsible Research Assessment,** a project initiated by RoRI in 2023.

AGORRA's outputs include the **RoRI Atlas of Assessment**, an online observatory which monitors national assessment and funding systems (informed by the typology in this paper); records changes over time; and provides a platform for assessment system design, experimentation and mutual learning among researchers, policymakers, research funders and

Our focus is on **national research assessment and funding systems,** defined as *"organised sets of procedures for assessing the merits of research undertaken in publicly funded organisations that are implemented on a regular basis, usually by state or state-delegated agencies"* (Whitley, 2007, 6).

Our typology and analysis do not include organisational procedures for the recruitment or promotion of researchers. or the assessment of research grant proposals by funding agencies. This work continues an **important tradition of comparative analysis of national research assessment and funding systems** (Geuna and Martin, 2003, Hicks, 2012, Debackere et al., 2018, Zacharewicz et al., 2019, Oschner et al., 2021, Sivertsen, 2023).

A recent **systematic review** of this literature (Thomas et al., 2020), which also included more than 300 opinion pieces, reveals that most contributions (including the review itself) implicitly assume that assessment and funding are always combined in national systems. An often-used term is therefore **performance-based research funding systems (PRFS**), which can be defined as *"national systems of research output evaluation used to distribute research funding to universities"* (Hicks, 2012, 260).

Whitley's broader definition, quoted above, is important. Information provided by international partners in AGORRA reinforced that:

- **Funding can be formally detached from research assessment and vice versa**, and there appears to be a trend in this direction.

- **Some institutional funding systems are not built on research evaluation but on indicators** representing already-performed assessments in other contexts – such as external funding and peer-reviewed publications.
- Some national assessment systems have **purposes other than funding allocation**, including accountability, organisational learning and strategic development.
- There is increasing interest in **assessing organisations and their procedures**, not only their outputs.
- Some systems include the **assessment of individual researchers**, rather than only their organisation's research.
- **National assessment systems may operate at multiple levels:** institutions, departments, research groups, and individual researchers – with different implications at each of these levels.

In contrast with other comparative accounts of performance-based funding systems, our typology and analysis also capture **national *ex post* systems** where, for instance, periodic evaluations of research performance are used to provide strategic advice.

Besides these differences, our criteria for inclusion of national assessment systems overlap with those employed in other comparative accounts, namely[1]:

- **Research must be the object of assessment**. Evaluations focusing only on the quality of degree programmes or teaching are excluded.
- **Research assessment must be *ex post.*** Evaluations of research grant proposals for project or programme funding are *ex ante* evaluations, and are excluded.
- **Research outputs must be evaluated in some way.** Systems that allocate funding based only on external research funding or PhD enrolment numbers are excluded.
- **It must be a national system.** University evaluations of their own research standing, even if used to inform internal funding distribution, are excluded.

This study also extends the literature comparing national systems by being **the first to incorporate a dynamic, longitudinal perspective**. We were motivated by t**hree research questions** in developing this new typology, and analysing changes over the past fifteen years:

- What **characteristics differentiate national research assessment and funding systems**, and how can these be categorized and compared?
- What **patterns of change** can be observed across national research assessment and funding systems over the period 2010-2024?
- How might **agendas of assessment reform play out over the next 5-10 years**?

[1] Adapted from Hicks, 2012, 252

As in earlier comparative studies (e.g. Hicks 2012; Kolarz et al. 2019; Zacharewicz et al. 2019), we employ an **analytical review method**, whereby our typology was **co-developed with country-specific experts** (a mix of academic researchers and senior staff within the funding and evaluation agencies involved in AGORRA). These experts first provided information via country-level templates, then offered follow-up feedback and clarifications through iterative exchanges, helping to apply the typology to their respective national systems. They also documented recent and ongoing changes to their systems, contributed to the final draft of this working paper, and provided input to the Atlas of Assessment observatory alongside.

Our analysis examines **thirteen countries'** national research assessment and funding systems: **Argentina, Australia, Brazil, Chile, China, Colombia, India, Italy, Mexico, Netherlands, Norway, Poland, and the United Kingdom (UK).**

This sample includes **seven countries from the Global South**, often under-represented in studies of national research assessment systems. The literature in this area is weighted towards peer-review led, **organisation-level evaluations**, more common in European and OECD countries – and less on **individual-level systems**, more prevalent in Latin America and other regions (Vasen et al, 2023). By including less-studied systems, including India, China, and five Latin American countries, our study attempts to capture key characteristics of other under-mapped examples. The **Appendix** to this paper contains a **concise summary of the 13 countries' systems**.

In the next section, we **introduce our typology** as it applies to the thirteen countries. We then present a **theoretical framework** for observing changes over time, by distinguishing between **three performance paradigms** in the development of research assessment and funding over the last forty years. Drawing on our typology and theoretical framework, we then analyse the **thirteen national systems**. We end by charting potential **future trajectories** for research assessment over the next five-to-ten years.

The international journal of science / 18 May 2023

nature

An opportunity to reassess research evaluation

Australia's overhaul of how a huge science funder measures research quality offers a chance to review what makes an assessment system work for everyone.

Research evaluation at the Australian Research Council (ARC), one of the country's main funding bodies, is set to get a makeover. Last month, an independent expert review recommended that the ARC scrap its 13-year-old research-scoring system, known as Excellence in Research for Australia (ERA), and its companion exercise, the Engagement and Impact assessment, which grades the real-world benefits of institutions' work. Both had been on hold since last August, awaiting the findings of the review.

This is a rare case in which an evaluation system can be rewritten from scratch. The ARC should take this opportunity to improve how it measures and communicates the

"Pitting institutions against each other encourages them to

against each other encourages them to poach academics and duplicate expertise, rather than fostering a collaborative, cross-disciplinary research culture.

The review wisely resists suggesting that the ERA be replaced with a purely metrics-based exercise to reduce overheads. Using metrics such as citations and journal impact factors is problematic, to say the least. For one thing, they do not capture replication studies and meta-analyses, which journals are increasingly publishing. Furthermore, studies in the social sciences and humanities – more frequently published in national, rather than international publications – are especially disadvantaged by such measures. And just counting citations in well-known journals doesn't do justice to other important work, such as building databases and software, and engaging with the public.

Ten years ago this month, the San Francisco Declaration on Research Assessment (DORA) enshrined the principle that evaluations should not rely too heavily on journal impact factors as a measure of research quality. The declaration, developed during the 2012 Annual Meeting of the American Society for Cell Biology, urges, among other things, that journal-based metrics not be used as a surrogate measure of research quality for assessing an individual scientist's contributions, or in hiring, promotion or funding decisions. More than 20,000 individuals and 2,800 institutions across 160 countries have signed DORA so far.

More has happened in the decade since. The 2015 Leiden Manifesto recommends that assessors anticipate how

A 2023 Nature editorial calls for more creative, evidence-informed approaches to the design of assessment systems.

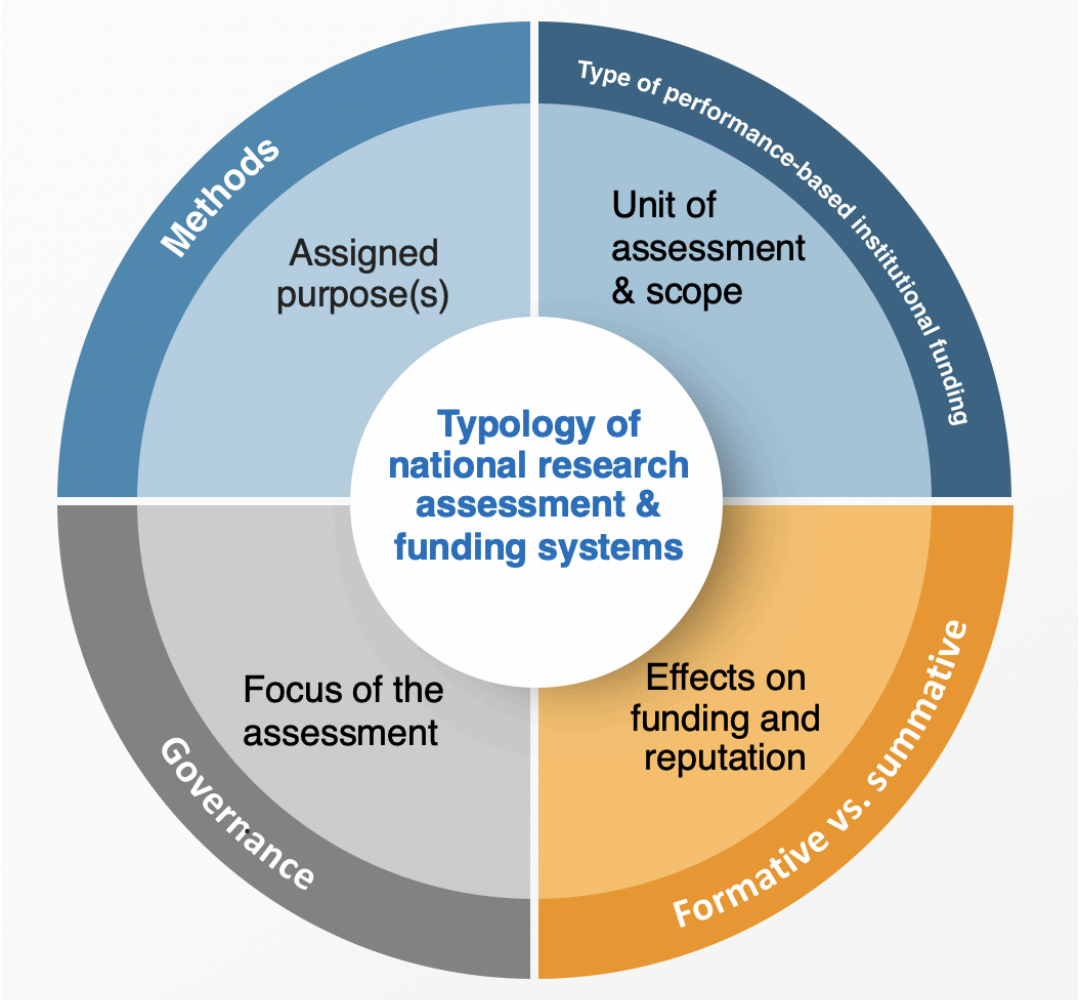
Through this analysis, we hope to **expand the shared language, menu of concepts and design tools** available to make sense of a diverse and dynamic landscape, and the various transformations now underway in many national assessment systems.

We hope our typology will also be a **useful resource for broader global discussions about reform of research assessment:** for example through CoARA (Coalition for Advancing Research Assessment) and DORA (Declaration on Research Assessment), where much focus to date has been on institutional-level frameworks (within universities or research funders), and less on national-level frameworks, as they operate in certain countries.

2. Typology of national research assessment & funding systems

Our typology is separated into **four core aspects**, which reflect **significant differences** in these systems, and **four subsidiary aspects**, which enable a **deeper understanding of their inner logics and effects**. Contemporary examples from our evidence base of national systems are used to illustrate each aspect of the typology. Our picture is further complicated by the fact that **national systems may combine different assessment types**, motivated by different purposes, and applied at different levels of aggregation. A country may therefore appear to exemplify one aspect more than once.

Figure 1: *Typology of national research assessment and funding systems, showing core aspects (inner circle) and subsidiary aspects (outer circle)*



The **four core aspects** are:

i) Assigned purpose(s)

Systems differ according to their assigned purpose(s) – usually one or more of the following:

- Funding allocation;
- Benchmarking and reputation;

- Accountability to governments or wider society;
- Organisational learning and strategic development;
- Statistics and mapping of research activity;
- Promotion of individual researchers;
- Accreditation

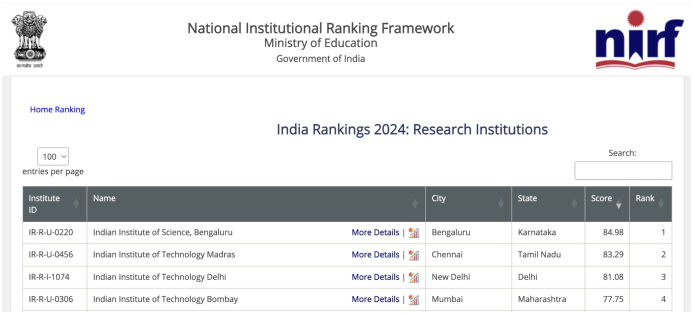
Where purposes are combined, the relative importance of each can be weighted differently by a given system.

Italy's Research Quality Evaluation (VQR), which is conducted by ANVUR (National Agency for the Evaluation of Universities and Research Institutes).

The idea of holding research performers accountable for their use of public funds and the results of research – is a frequently recurring purpose across our sample of systems. Many countries' systems combine accountability with other objectives like funding allocation and reputation building (funding allocation and reputation). Some, like the **UK's** Research Excellence Framework (REF) or **Italy's** Research Quality Evaluation (VQR), directly link evaluation results to funding outcomes; while others, such as **Chile's** National Accreditation Commission place more emphasis on the reputational consequences of research evaluation results.

This combination of purposes extends to national systems for individual assessment and promotion in **Argentina, China, Colombia** and **Mexico**. In Latin America, such systems routinely assess *"the individual performance of academics based on their academic activities and outputs and assigns them a 'researcher category', which carries prestige and, in many cases, additional monthly rewards"* (Vasen et al, 2023, 244).

India's NIRF (National Institutional Ranking Framework)



Indicator-based systems, like those of **Norway** and **India**, also incorporate funding and reputational elements to varying degrees, with **India's** NIRF (National Institutional Ranking Framework) characterized by low reliance on funding and high emphasis on reputation). By contrast, a smaller number of systems, exemplified by the **Dutch** Strategy Evaluation Protocol, **Argentina's** PEI (Programa de Evaluación Institucional) and **Norway's** Disciplinary Evaluation, prioritize organisational learning and strategic development (alongside accountability in the Dutch case).

ii) Unit of assessment and scope

Systems may collect information about and assess:

A. *Disciplines across an institution;*
B. *The institution as a whole;*
C. *Units or research groups within the institution;*
D. *Individual researchers.*

The scope for an assessment may include all eligible staff within the unit of assessment, or a certain selection of them. A unit of assessment is often a university department, which may also represent a discipline (though many departments will combine a range of disciplines).

The UK system allocates just over £2 billion per year of research funding to institutions as a result of the REF.

The level of assessment may differ from the level at which funding is then allocated. Both need to be categorized. As an example, the **UK** REF assesses the quality, impacts and research environment of up to 34 units of assessment within institutions, but the outcome is then reflected in a multi-year funding allocation to the institution as a whole.

More than one unit may be addressed in the same system. **Italy** combines A) and B). Norway combines A), B) and C) in its disciplinary evaluation. In **China**, the word 'double' in the so-called 'Double First-Class Evaluation' signals a combination of university and disciplinary-level assessments. Finally, some countries e.g. **Argentina, China, Colombia** and **Norway.** have more than one framework, each targeting different purposes and units of assessment

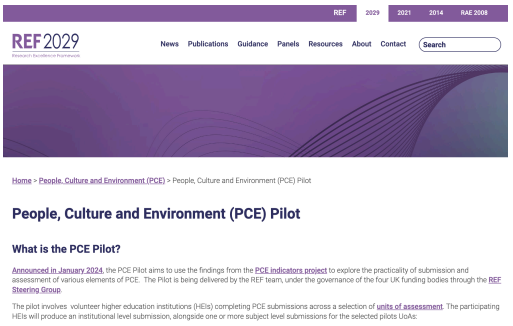
iii) Focus of the assessment

Systems may focus and collect evidence on different aspects of research performance:

A. *Scholarly outputs;*
B. *Scientific significance and impact (citations);*
C. *Societal impacts (collaboration and co-creation, business R&D, policy influence, public engagement, spin-outs, technology transfer);*
D. *Competitive grants;*
E. *Organisational performance;*
F. *Research cultures and environments;*
G. *Performance of individuals;*
H. *Other*

The **UK's** assessment system, first introduced in 1986, initially focused on scholarly outputs and environment. Societal impacts (via impact cases) were added in the two most recent exercises (with a weighting of 20% in 2014, then 25% of the 2021 exercise). Organisational

performance (as part of research environment) was included in 2021; and a pilot is now underway of an expanded framework and indicator set for “people, culture and environment”, to be included in REF 2029, with a notional allocation (pending results of that pilot) of 25% of the exercise.



REF 2029 ‘People Culture and Environment (PCE)’ Pilot is now underway.

Revenues from competitive grants are often an indicator of organisational performance in indicator-based systems. **Norway** and **Poland** are two examples – while **Argentina, China, Colombia** and **Mexico** all have specific systems for assessing the performance of individuals.

iv) Effects on funding and reputation

Systems are often debated in terms of their intended and unintended effects. Dahler-Larsen (2014) points out that systems always have ‘constitutive effects’ simply by being implemented. Through their official aims and practical operation, systems may determine:

A. *Funding and reputation*
B. *Only reputation*
C. *Other significant effects (e.g. strategic development, accreditation)*

The relative importance of each of these – and the strength of influence exerted on funding and reputation – can be weighted (strong-medium-weak) by users of the typology.

Many national systems have some sort of connection to both funding and reputation. Examples where assessments ostensibly only influence reputation are **Chile, Colombia** (2 of 3 systems), **India**, **the Netherlands**, and **Norway** (1 of 2 systems). But reputation may still indirectly influence funding. The ERA (Excellence in Research for Australia) system initially influenced both funding and reputation. After the funding link was removed, the reputational effects were still considered important by Australian universities.

Aside from funding and reputation, other significant effects may be felt in strategic development and learning at an institutional, local or national level. This is the case in **Chile, Colombia** (1 of 3 systems), **the Netherlands** and **Norway** (1 of 2 systems). Systems with a strong influence on funding and reputation will generally influence strategic development in other ways. Individual level assessments – as found in **Argentina, Brazil, China, Colombia** and **Mexico** – may influence career prospects, salary levels, and the availability of resources for performing research.

Polish Accreditation Committee

A further possible effect is accreditation: the outcome of an assessment may influence the right of an institution to grant certain degrees; provide certain courses, or establish new professorships. **Poland** is one example.

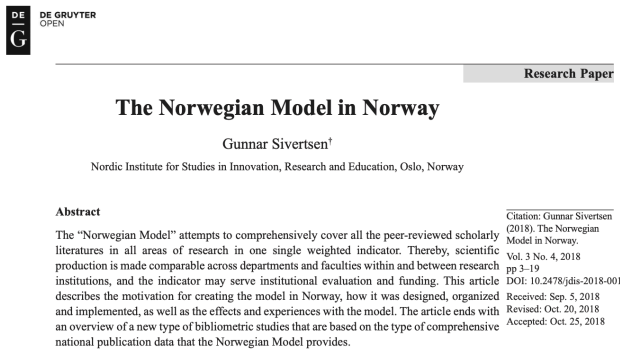
The **four subsidiary aspects** of the typology are:

v) Methods

National systems often rely on **expert panels** informed by **peer review, national statistics** and other **scientometric indicators**, but may differ in how peer-review and/or expert advice are organised, and to what degree an assessment is shaped by data or more deliberative modes of judgement and evaluation. **Types of indicators** and **underpinning data sources** may also vary (e.g. through a reliance on a commercial database, or an open system like OpenAlex). The relative importance of each method can be weighted.

The balance between **quantitative performance indicators and qualitative peer review** remains a persistent tension in assessment and funding systems. Many systems aspire to operate according to principles of **'informed peer review'** whereby quantitative indicators support, but do not replace, expert-led assessments (Butler, 2007; Wilsdon et al., 2015).

Some systems apply different methods by field: in **Australia's** ERA and **Italy's** VQR, social sciences and humanities were assessed using peer review, while STEM panels were more likely to use bibliometrics. ERA has now been paused pending the introduction of **Australia's** new National Research Evaluation and Impact Framework (NREIF); and in the current round of VQR, the role of bibliometrics has been reduced. The **UK** REF allows expert panels discretion in deciding whether to make any use of citation data, but expert review remains the primary method.



DE GRUYTER OPEN

Research Paper

The Norwegian Model in Norway

Gunnar Sivertsen†

Nordic Institute for Studies in Innovation, Research and Education, Oslo, Norway

Abstract

The "Norwegian Model" attempts to comprehensively cover all the peer-reviewed scholarly literatures in all areas of research in one single weighted indicator. Thereby, scientific production is made comparable across departments and faculties within and between research institutions, and the indicator may serve institutional evaluation and funding. This article describes the motivation for creating the model in Norway, how it was designed, organized and implemented, as well as the effects and experiences with the model. The article ends with an overview of a new type of bibliometric studies that are based on the type of comprehensive national publication data that the Norwegian Model provides.

Citation: Gunnar Sivertsen (2018). The Norwegian Model in Norway.
Vol. 3 No. 4, 2018
pp 3–19
DOI: 10.2478/jdis-2018-001
Received: Sep. 5, 2018
Revised: Oct. 20, 2018
Accepted: Oct. 25, 2018

The Norwegian Model

Indicator-based systems offer standardized approaches, exemplified by the **Norwegian** Indicator, which serves as a nationwide system for allocating resources based in part on a comprehensive publication database, a publication indicator, and a performance-based funding model (Sivertsen, 2018).

Similarly, **Poland's** Parametric Exercise (EJDN) aggregates points on a range of quantitative performance criteria, including bibliometrics, PhD graduations, and educational indicators, which inform the distribution of core funding, and may affect accreditations.

The use of algorithms and formulas is not limited to indicator-based systems. In **Italy's** VQR, panel scores given to units of assessment feed into competitive performance rankings and formula-based distribution of research funding.

Given their primary assigned purposes of generating strategic advice to research performing units, indicator-informed peer review-based evaluations like the **Dutch** Strategy Evaluation Protocol and the **Norwegian** disciplinary evaluations, work differently. Expert panels use peer review to deliver narrative reports on performance and ongoing organisational improvements.

Pressures to ensure appropriate uses of quantitative indicators have grown in recent years with the rise of **'responsible metrics'** as a trans-national professional reform movement across research policy and universities (Rushforth and Hammerfelt, 2023) (see Section 3. below).

vi) Type of performance-based institutional funding

Systems that affect institutional funding directly may appear in three main types:

A. evaluation-based funding (the use of peer review and expert panels),
B. indicator-based funding (direct use of performance indicators),
C. and funding contingent on performance agreements between the funder and the individual research organisations.

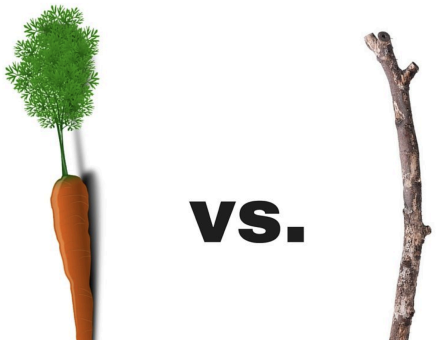
Among the thirteen countries included in this overview only **Norway** and **Poland** have indicator-based organisational funding systems, and none have organisational funding based on performance agreements. These types are however more widespread than our sample indicates (Sivertsen, 2023).

Indicators may also have an important role in informing evaluation-based funding systems. This is the case in **Argentina, Chile, China, Italy** and **Mexico**. Expert panels in **Australia's** ERA used to draw heavily on supporting contextual indicators in select disciplines (mainly STEM).

vii) Formative versus summative

Different purposes may result in different **guiding orientations** for an assessment.

A. A *formative* evaluation learns from the past and looks forward, supporting organisational learning and development, and incentivising change
B. A *summative* evaluation looks at past performance, audits whether goals or expectations have been reached, and serves decisions and/or resource allocation.



A more colloquial version of this distinction is that drawn between **carrots (formative)** and **sticks (summative)**. Where these orientations are blended or combined, the relative importance of each can be weighted using expert judgement.

Systems tend to be summative if they rely mainly on empirical evidence for the assessment of past performance. Also, if they determine funding, the outcome of the assessment has to be translated into a quantitative **funding formula.**

Of our thirteen systems, only **Chile, the Netherlands**, one of **Australia's** (now ceased) systems, and one of **Norway's** two systems, can be said to have a summative orientation. But several others combine summative and formative elements, particularly when the effect on funding and reputation is weak (Sivertsen, 2023). An example is the three purposes of the **UK's** REF for its next cycle in 2029:

"1) Inform the allocation of block-grant research funding to HEIs based on research quality; 2) provide accountability for public investment in research and produce evidence of the benefits of this investment; 3) provide insights into the health of research in HEIs in the UK" (REF, 2029)

The first purpose here is largely summative, and the third formative, while the second reflects both orientations. So this can be a question of emphasis rather than a binary distinction.

viii) Governance

Governance is about how assessment systems are designed, implemented, and organised on a continuing basis with distributed responsibilities. Systems may differ according to:

A. The role of the agency responsible and its relations to government on one side, and research organisations on the other. (Systems tend to be under formal control of a central government agency or arms-length body like a research council);
B. The involvement of and collaboration with the national academic community in the design, implementation, management, and evaluation of the system;
C. The transparency and predictability of the methods and results (high-medium-low);
D. Mandatory vs. voluntary vs. incentivized participation.

Research assessment and funding systems develop within specific national and historical contexts, with varying degrees of political and administrative centralization. The role of agencies responsible for these systems differs across countries, particularly in their relationship to government and the academic community.

Some systems are directly controlled by government agencies, such as in **Poland, India** and **Colombia**, where ministries of science, education or equivalent administer assessments. But a trend over the past three decades has seen the growth of intermediary organisations like funding agencies and research councils (Braun, 1998, Braun and Guston, 2003) overseeing ex-post assessment and funding, as seen now in **Brazil, Norway, UK**, and the **Netherlands.**

Timetable for piloting & consultation for REF 2029 (extract)[2]

The extent of academic community involvement in system design also varies significantly between nations. Countries like the **UK** and the **Netherlands** regularly consult with academics, research managers, and sector groups when making periodic adjustments to their systems. Others, such as **Colombia** and **India**, offer limited opportunities for input.

Within countries, levels of academic consultation can differ across systems. The presence or absence of meta-evaluations of assessment and funding systems is another indicator of accountability to the research community and other stakeholders. At present, few

[2] https://2029.ref.ac.uk/about/timetable/

countries routinely commission independent meta-evaluations and make them publicly available. The **UK's** REF (Stern, 2016, Digital Science, 2015) and **Italian VQR** (Expert Review Panel, n.d., Galderisi et al., 2019) have periodically had meta-evaluation reviews commissioned.

The transparency and predictability of methods and results in national assessment systems are variously influenced by the choice of method, purpose of assessment, and the consequences of results for funding and reputation. **Indicator-based systems**, due to their standardized nature, theoretically offer **higher reproducibility and transparency**, as exemplified by the **Norwegian** Indicator.

However, this principle is not universally upheld. **Peer review-based evaluations,** being less standardized, generally offer **lower reproducibility and transparency**. The purpose of the evaluation plays a role, as evaluations linked to funding may intentionally limit transparency due to potential litigation risks, as seen in the **UK** REF's closed-door panel discussions, where panel scoring of individual research outputs are destroyed and not made public, to uphold confidentiality and data protection.

Peer review evaluations oriented towards delivering narrative reports and strategic advice, like the **Dutch** Strategy Evaluation Protocol (SEP), exhibits low transparency in how results are produced and used. A distinction can be made here between the public availability of the overall results, versus how, and to what degree, feedback is utilized by the evaluated units (Gras, 2022).

Finally, the extent to which assessment exercises are **mandatory, incentivized or voluntary** varies across national assessment systems. **Brazil's** CAPES, the **Dutch** SEP and **Italy's** VQR are examples of systems for which participation has effectively become mandatory through regulation or laws.

Systems in which research performing organisations or researchers are **incentivized to participate but not legally mandated** (through the 'carrot' of financial or reputational rewards or potential negative costs of not participating), include **Argentina's** PRINUAR, the **UK's** REF, and **China's** national assessment and selection systems for elite individual researchers. Other systems where participation is incentivized, include **India's** NIRF, **Chile's** CNA, **Colombia's** High-Quality Accreditation Model, and **Mexico's** SNI.

In **nations with multiple assessment systems**, certain systems may be mandatory, while others are incentivized or voluntary. For example, the **Norwegian** indicator-based funding system for research is mandatory for universities until 2025 (and still mandatory in hospitals and public research institutes beyond that), but subject-specific evaluations are only normatively expected (so in our typology, count as voluntary); and in **China**, the National Disciplinary Evaluation is voluntary, but the Double-First Class Evaluation is incentivized.

3. Making sense of ongoing developments in national research assessment systems

Our multi-case, thirteen-country comparison of the development of national assessment systems from 2010 to 2024, combined with insights from earlier comparative studies, reveals a pattern: that the purposes, foci and methods of research assessment, have changed and continue to evolve, to varying degrees over the last four decades. These changes can be mapped according to **three broad paradigms of research performance** popularized within research policy at various points in time:

1. The paradigm of *professional-disciplinary evaluation*
2. The paradigm of *excellence and competition*
3. The paradigm of *responsible research assessment*

Originating with Thomas Kuhn (1962), and adapted for policy analysis by Peter Hall (1993), the term **paradigm** can be defined as *"a shared model of reality that guides policy makers' problem-solving activities"* (Carson et al., 2009, 18).

Crucially, one paradigm does not straightforwardly replace another: in practice, there is often a **layering** (Aagaard, 2017, Capano, 2023) of different paradigmatic ideas, which emerge and evolve in context specific ways, and shape assessment systems differently over time.

In labelling our paradigms with terms such as ***excellence*** and ***responsible research assessment***, we acknowledge giving them normative **'member categories'** used by others in the field. We consider such labels however to be useful analytical tools for interpreting broad developments in assessment systems. Such labels typically emerge once developments are underway, serving to capture a sense of transformation within research assessment. At the same time, these labels are also **performative**: they may be used to assist some actors in agenda-building, or to justify, validate, or strengthen certain positions and decisions (Rip, 2000).

Let us outline the three paradigms in more detail:

Paradigm 1: Professional-disciplinary evaluation

The first paradigm of research performance, which became dominant before the advent of formalized, external assessment, was driven by the **professional-disciplinary tradition** (Elzinga, 2012). Here academic departments operated with relatively high levels of autonomy, relying on internal disciplinary standards to guide decision-making around hiring, promotion, and internal allocation of resources (Whitley, 2000).

Through this period, from Vannevar Bush's 'endless frontier' onwards (Bush, 1945), a great deal was written about changes in the policy rationales, design, and methods of national public science systems. In 1994, physicist John Ziman described the end of an era of 'big bang' investment in public sciences that had defined the post-war period in many Western countries (Ziman, 1994). For Ziman, these countries were now entering a 'dynamic steady state' era of science policymaking, defined by moderate increases, stabilization, or cuts to

research funding, with economic crises, ideological shifts, and changing cultural conceptions of S&T making politicians more reluctant to sustain levels of funding enjoyed in the rapidly expansionist post-WWII period.

For Ziman, the clearest manifestation of this newly incrementalist regime was the then-recent emergence of the world's first periodic ex-post national assessment and funding system, the **Research Assessment Exercise (RAE)** in the UK. Accompanied by decreases in block grant funding in favour of project funding, the RAE promised rationally to support the allocation of dwindling funds, provide research quality assurance mechanisms, ensure accountability for scarce public funds, and promote the UK's global competitiveness in research.

Figure 2: *The eight cycles of UK research assessment since 1986 (Curry, Gadd & Wilsdon, 2022)*

Date	Exercise	Coordinating body	Key features
1986	Research Selectivity Exercise	Universities Grants Committee	37 cost-centres; 4-part questionnaire on research income, expenditure, planning priorities & output
1989	Research Selectivity Exercise	Universities Funding Council	152 units of assessment; 70 peer review panels; 2 outputs per member of staff
1992	Research Assessment Exercise (RAE)	HEFCE	HEIs select which staff to submit; 5-point scale; 2800 submissions to 72 UoAs; introduction of census date
1996	Research Assessment Exercise (RAE)	HEFCE	Up to four outputs per researcher; 69 UoAs
2001	Research Assessment Exercise (RAE)	HEFCE	2600 submissions to 69 units of assessment; 5 umbrella groups of panel chairs for consistency
2008	Research Assessment Exercise (RAE)	HEFCE	67 sub-panels under 15 main panels; results presented as quality profiles
2014	Research Excellence Framework (REF)	HEFCE	4 main panels; 36 sub-panels; introduction of 20% impact element
2021	Research Excellence Framework (REF)	UKRI (Research England + devolved funding councils)	All staff with significant responsibility for research included. Impact 25% weighting. Flexible number of outputs.

Even before the first version of the UK's exercise was conducted in 1986, national research policymakers in the UK and elsewhere became increasingly concerned with the competitiveness and economic performance of **national 'knowledge economies',** influenced in part by comparative international evaluations by organisations like the OECD, World Bank, and European Union (Wagner et al., 2001, Henriques and Larédo, 2013, Lepori et al., 2007). A belief that increased investment would improve the overall quality of research and the economic and societal benefits that flow from it, was giving way to a view that funds should be targeted to promising areas of S&T, or concentrated in leading institutions.

This period was particularly advantageous for biomedical and health sciences, with other domains (particularly the social sciences and humanities) experiencing more uneven investment and cuts (Jones and Wilsdon, 2018). In many national systems, a new **performance paradigm focused on excellence and competition** was taking hold in research policy, influenced by the frameworks of new public management that were simultaneously coming to dominate other areas of policy and public spending. Yet despite these developments, the

1970s and 1980s were for many academics an era where research assessment was dominated by disciplinary standards in individual fields, rather than explicit consideration of interdisciplinary, policy or societal needs (Elzinga, 2012, Whitley, 2000).

Paradigm 2: Excellence and competition

From the early 1990s onwards, a performance paradigm emerged with a **strong focus on 'excellence'** in research assessment and funding. As one symbol of this, the UK's RAE would eventually (from 2009) change its name to the **Research Excellence Framework (REF)**. In many countries, excellence became a common denominator that could be applied in all areas of research (Jong et al, 2020). It was also a strong instrument of prioritization: we will only fund the best, and the funded will get more (Scholten et al., 2021).

This focus on measuring research performance coincided with the **increasing availability of bibliometric data** and its usage in performance-based funding systems. The concept of excellence was initially seen as a unifying standard. But in practice, such approaches were poorly aligned with the research and evaluation traditions of certain fields, including much of social sciences and humanities.

The rise of the 'excellence paradigm' was the focus of an earlier RoRI project – see e.g. this working paper.

The excellence paradigm has been attributed to wider currents of globalization, and promises of increased efficiency through competition (Elzinga 2012, Hicks 2012). **Periodic assessments, performance-based funding, use of quantitative performance indicators, and incorporation of public policy goals** into evaluation criteria are all features of this paradigm (Rip, 2004, Whitley, 2019). Another example of the rise of this paradigm is found in the stated policy goal at a European level to establish *'more effective national research systems – including increased competition within national borders and sustained or greater investment in research'* (European Commission, 2012).

As implied by the notion of 'layering', **the excellence paradigm does not replace the disciplinary paradigm per se, but builds on and reconfigures it**. Thus, systems that promote accountability and competition, including those with the word 'excellence' in their title (like Excellence in Research Australia and the UK REF), have continued to uphold certain practices synonymous with the previous paradigm (e.g. appointing peer review panels to administer evaluations along disciplinary lines). Likewise, systems that utilize bibliometrics and rankings often transform peer reviewed scholarly outputs into quantifiable data points for aggregation.

Paradigm 3: Responsible research assessment (RRA)

More recently, from 2012 onwards, the performance paradigm as manifest in national research assessment systems has entered a third stage in some countries, with the emergence of the **'responsible research assessment' (RRA) agenda** (Curry et al., 2020, Curry et al., 2022, IAP-GYA-ISC, 2023, Benamara et al., 2024). This can be read as a response to perceived 'public value failures' in existing paradigms of research assessment (Bozeman and

Sarewitz, 2011), with growing discontent towards both the traditional professional disciplinary evaluation paradigm (for excessive self-referentiality, specialization and overemphasis on scientific publications and citation impact indicators) and the excellence paradigm (for excessive emphasis on competition and selectivity).

A shift to this third paradigm is well reflected in this 2023 report, co-published by three global scientific bodies.

Critics argue that excellence-oriented national assessment and funding systems are creating **perverse consequences in research**, such as hyper-competition and poor academic work cultures, task reduction, goal displacement, and deterring inter- and trans-disciplinarity (de Rijcke et al., 2016).

RRA attempts to move the focus of assessment towards promoting a **broader range of priorities**, including: equity, diversity and inclusion (EDI); improving research culture; open science; team science and other forms of collaboration; and addressing societal challenges and missions. The convergence of various science reform movements from the mid-2010s onwards has generated more visible, organised momentum for assessment reform (Rushforth and Hammarfelt, 2023).

CoARA's vision and mission statement.

New initiatives have provided visibility and platforms for engagement with this agenda. The most ambitious initiative, the **Coalition for Advancing Research Assessment (CoARA)** now involves over 800 research funding organisations, research performing organisations, and assessment authorities. It builds on calls for responsible uses of quantitative indicators, advanced in the San Francisco Declaration on Research Assessment (DORA, 2013), Leiden Manifesto (Hicks et al., 2015), and The Metric Tide (Wilsdon, 2016).

Similarly, the **Global Research Council (GRC)** working group on research assessment recently set out a framework of **eleven dimensions of RRA** (Benamara et al., 2024). CoARA and the GRC incorporate but go beyond calls for responsible metrics, by outlining other marginalized features and qualities of research they want to see rewarded and recognized in various levels of research assessments, including professional service work, open science contributions, research integrity, equity and inclusion, and societal engagement.

So the **RRA performance paradigm** combines existing and emerging ideas, practices, and criteria, seen as alternatives to the shortcomings of the disciplinary and excellence paradigms – even if not all the practices or ideas are new in themselves. Twelve years after the European Commission (2012) published its guidelines for *A Reinforced European Research Area Partnership for Excellence and Growth*, a high watermark of the excellence paradigm, in 2024, the EC published an **Action Plan to implement the Agreement on Reforming Research Assessment**, in collaboration with CoARA (European Commission, 2024). This plan suggests that research assessment has now entered the early stages of a third paradigm in which broader criteria for research assessment are promoted.

But the **consensus in support of this third paradigm remains fragile**, and the focus of resistance from interests within academia, governments and wider society, more vested in one of the earlier two paradigms. Ripple effects of the second Trump presidency in the US on these debates are also perceptible, as part of a broader pushback on EDI agendas – what some have dubbed the 'Great Vibe Shift' (Leslie, 2025).

Can national systems be expected to change at the same time and in the same direction?

CoARA is presently an example of how research performing and funding organisations may collaborate across nations to reform the criteria and procedures for research assessment in the same direction. However, **CoARA mainly focuses on individual-level assessments by universities and funders** for recruitment, promotion and external project funding.

CoARA's Agreement on Reform of Research Assessment does not approach national assessment and funding systems with the same clarity (Sivertsen and Rushforth, 2024). This may be because national systems of assessment are largely shaped by governmental agendas and led by organisations which aren't themselves signatories to CoARA. **To what degree then can we expect countries to change their national assessment systems to reflect the rise of the RRA performance paradigm?**

Based on experience with the design, implementation, development and discussions of performance-based organisational funding systems in 26 countries, Sivertsen (2023, 90) explains why systems may differ and change at different rates with different influences:

> *Although some systems may seem similar across countries, they are never the same and they are modified all the time. PBFS [performance-based funding systems] differ because they are anchored in the local traditions and mechanisms of state budgeting and embedded in the local negotiations about priorities and developments in the higher education sector. They are dynamic because they are continuously contested and thereby often adjusted.*
>
> *Countries also mutually learn from each other and inspire changes in their PBFS. The systems are conservative as well. Once implemented, they become games with rules and specific terminologies and infrastructures that are difficult to change. Also, they need to be predictable because they influence budgets and the spending of tax*

revenues on the funding side. There is a need to ensure some stability of budgets at the institutions.

A further complication is that even when countries are influenced by the same performance paradigm, *'the final result is not convergence but different interpretations of the same general recipe*' (Capano, 2023). Layering of the nascent third paradigm – RRA - plays out differently and to varying degrees in our sample of thirteen national systems.

Preliminary observations of an emerging RRA paradigm

Let us now highlight **critical junctures and patterns of change** across our sample of countries, from 2010 to 2024. In doing so, we argue: (a) different aspects of **these three paradigms are 'layered' onto one another** to greater or lesser degrees, meaning wholescale replacement of one by another does not occur (c.f. Aagaard, 2017, Capano, 2023), and (b) **the extent to which a third RRA paradigm has begun to affect national assessment systems** (or is even visible at all) varies considerably across our featured countries.

Early manifestations of this emerging third phase came with the introduction of societal contributions and impacts as a new criterion in **Australia** (Williams and Grant, 2018) and then in the **UK's** REF 2014 (Martin, 2011). The REF's adoption of impact in turn inspired other countries, such as the **Netherlands, Brazil,** and **Poland** to utilize similar criteria in a range of ways. The **Netherlands** and **Norway**, like the UK, incorporated societal impact or relevance into existing systems, and in **Poland**, societal impact criterion was added to the EJDN evaluation system, as one of many changes in the 2018 Law on Higher Education (Wróblewska, 2025).

Australia ran its **Engagement and Impact (EI)** framework as a separate system alongside the existing ERA (which continued to reward disciplinary-based excellence) – but this ran only once, in 2018. Societal impacts are the most prominent and well-integrated dimensions of RRA to feature in national assessment systems to date, with a growing number of systems accommodating some variation on this.

Other RRA components, like open science, multilingualism, and responsible metrics, have had varying levels of uptake across the systems in our sample. Though **Poland's** EJDN incorporated societal impact, it has so far not accommodated additional elements of RRA. Ongoing debates in **Australia** about a framework to succeed the ERA reflect only modest engagement with the RRA agenda – despite a recent report by Australia's now-former Chief Scientist, Cathy Foley, which sat firmly within the RRA performance paradigm (ACOLA 2023).

Quantitative, indicator-based systems and excellence criteria feature prominently in many **Latin American** assessment systems, but there are also visible efforts to diversify assessment approaches. Efforts to counter English-language bias and support publishing in Spanish and Portuguese have grown, by integrating regional indexes like SciELO, RedALyC, and Latindex into national systems (Beigel, 2025).

In 2014, **Argentina's** CONICET approved a resolution for the social science and humanities equating journals indexed in regional databases with those in international indexes like Web

of Science and Scopus. CONICET's researcher career system also has qualitative, narrative components integrated into its evaluations, alongside bibliometrics and interviews. As mentioned earlier, **Brazil's** 2021–2024 CAPES evaluation cycle also integrated 'impact of society' as an explicit criterion. These changes have been amplified by regional advocacy networks such as the **Latin American Council of Social Sciences (CLACSO)** and **Latin American Forum for Research Assessment (FOLEC),** whose 2022 principles sought to raise awareness, adapt, and extend global frameworks like DORA and the Helsinki Initiative on Multilingualism in a regional context (CLACSO, 2022).

Citation: LI Xiaoxuan, XU Fang. How to Break the "Siwei"?—Practice and Enlightenment Based on Research Institute Evaluation of Chinese Academy of Sciences [J]. Bulletin of Chinese Academy of Sciences, 2020 (12): 1431–1438.

How to Break the "Siwei"?—Practice and Enlightenment Based on Research Institute Evaluation of Chinese Academy of Sciences

LI Xiaoxuan[1,2], XU Fang[1,2]

1. Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190, China;
2. School of Public Policy and Management, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: In October 2018, five ministries and institutions, i.e., Ministry of Science and Technology, Ministry of Education, Ministry of Human Resources and Social Security, Chinese Academy of Sciences (CAS), and Chinese Academy of Engineering, collaboratively started the special action of breaking "Siwei," which means "Four-Only" problems, i.e., only papers, only titles, only education background, and only awards. Most researchers in universities and research institutions have both expectations and concerns. There are different opinions on how to break the "Siwei." On the basis of the analysis of the development of evaluation conducted by CAS for more than 20 years, this study holds the view that CAS has explored a way of breaking the "Siwei" and formed the CAS mode in research institute evaluation, which is expected to provide a case for reference on how to break the "Siwei."
DOI: 10.16418/j.issn.1000-3045.20201116002-en

Keywords: break the "Siwei"; Chinese Academy of Sciences; research institute evaluation; CAS mode; science evaluation

Over the years, science evaluation, particularly in the basic

indicators should be added. As affected by various opinions,

In 2018, five ministries and institutions in **China** issued a special action, calling on the Chinese research system to curb the dominance of **'the four onlys'** ('only papers, only titles, only educational background, and only awards') as the primary criteria for research evaluation and talent recognition across the system (Xiaoxuan and Fang, 2020). This coincided with critiques in national policies of 'Science Citation Index worship', echoing some of the language of DORA and the Leiden Manifesto (Zhang and Sivertsen, 2020).

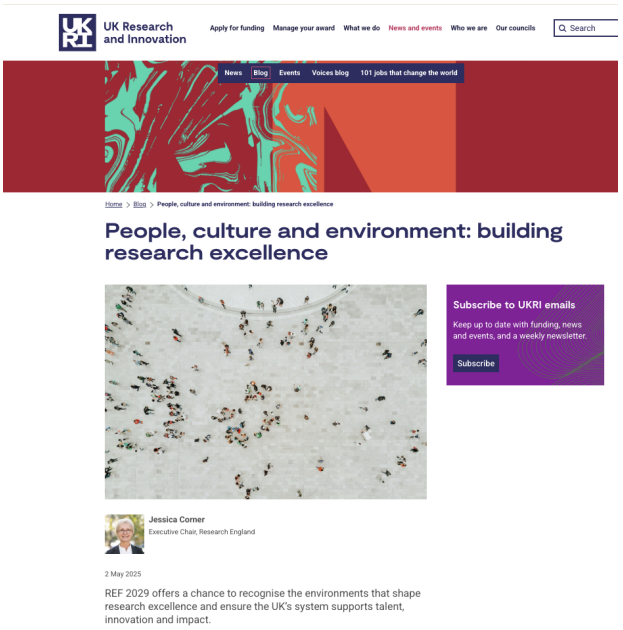
Even before the 2018 special action, some leading research institutes had already started moving away from purely quantitative, output-based definitions of academic performance, with the Chinese Academy of Sciences (CAS) among those taking a leading role in these reforms. From 2011 onward, the **CAS 'Research Institute Evaluation'** process was reformed into an achievement-oriented evaluation system primarily based on qualitative peer review, with quantitative indicators providing only supporting information. This shift was the culmination of longer-term efforts by CAS to increase the role of expert judgment and counter-balance negative effects of the metrics-dominated approach it had initially adopted in the early 1990s (Xiaoxuan and Fang, 2020).

One feature of the emerging RRA phase is a **shift in assessment methods from an exclusive focus on outputs** (heavily emphasized in Hicks's 2012 definition of PBRF systems) **to include more process-oriented indicators and narrative-based methods of research performance**. Of the performance based funding systems we studied, the UK's REF has experimented the most with moves away from traditional output and results-based forms of evaluation.

This is particularly evident in the introduction of 'institutional-level environment statements' in REF 2021, which continued a shift in emphasis towards **process-oriented indicators** that started in RAE 2008. The environment statement required institutions to submit strategic statements that focused on, for instance, support for inclusive research culture for staff and

research students, and steps to ensure sustainability of the unit through investing in people and infrastructure (Inglis et al., 2024).

The elevation of process and input-oriented indicators is a significant development for performance-based funding systems. Where process indicators were mentioned at all in earlier comparative studies, they were treated as unrelated to research performance (e.g. Hicks' 2012 paper explicitly positioned process indicators as separate from this type of system). The growing importance of process-oriented indicators over the past two cycles of the UK REF has gone largely under the radar of existing comparative studies.



This May 2025 article by Dame Jessica Corner, Executive Chair of Research England, summarises the case for PCE as a core strand of an effective assessment system.

This trend is now being intensified further via proposals for an expanded weighting of 25% for the **'People, Culture, and Environment' pillar of the next REF** (REF 2029) and an ongoing pilot process to develop and test new indicators. Earlier accounts that saw process indicators of research management and infrastructure as falling outside the scope of research assessment systems reflected prevailing views at the time, but the emerging paradigm of assessment treats **research culture and process indicators** as increasingly important in assessing research performance.

Open science, itself a diverse umbrella term (UNESCO, 2021), has begun to feature as a requirement in some national assessment systems. The **UK's REF 2021** included an open access mandate as a condition for eligibility of certain output types, as has the most recent version of the **Italian VQR 2020-2024**. Not all our thirteen countries have such requirements – and within the period covered, none had yet committed to **using open information as the basis for evaluations** of performance in their national assessment systems, as advocated by the **Barcelona Declaration on Open Research Information** (Barcelona Declaration 2024).

Elsewhere under the open science umbrella, the **Italian VQR 2024's assessment** is based on three criteria: originality, methodology, and the impact of the submitted research products. Unlike past VQR exercises, the methodology has been re-defined to encourage evaluators to pay attention not only to the rigor of research steps but also to **reproducibility, transparency, accessibility, and the reuse of data** (whether used) in the publications.

This is reflective of national assessment systems' gradual responsiveness towards pressures by science reform movements on **agendas of open science, research integrity, and metascience**. In Latin America, SciELO, RedALyC, and Latindex's infrastructures have long supported integration of Diamond Open Access journals (free to read and publish in), facilitating their inclusion in formal research assessment processes. These indexing systems

have also implemented quality control mechanisms to screen so-called predatory or spurious journals (Beigel, 2025). Open access publishing was also prioritized in **Colombia's** 2024 Research Groups Assessment Model.

Agencies responsible for several national assessment systems have signed **declarations such as DORA**, to commit to responsible uses of (particularly quantitative) research indicators. These include **NWO** and **KNAW** (overseeing the Dutch SEP) and **UKRI** (of which **Research England**, one of the four UK funding bodies who govern the REF, is part). **ANVUR** (overseeing Italy's VQR) is also a signatory to CoARA, and in 2023, **Chile's National Research and Development Agency (ANID)** became the first Latin American institution to sign CoARA.

ANID in Chile signs the CoARA agreement 2023

In most countries, no single agency oversees all research assessments (including ex-ante and ex-post), so voluntary initiatives like DORA and CoARA can be signed by different actors in a given national context. In **Argentina**, for instance, CONICET signed DORA in 2022, but is the only national agency (responsible for one strand of the assessment system) to have done so.

Individual universities may also sign these agreements, without the national assessment agency doing so, as in **Poland**. It is unclear how multiple actors signing these voluntary action initiatives may translate into concrete reforms of research assessment practices at a national system level, especially as the initiatives themselves have only weak enforcement or compliance mechanisms. There are other types of assessment beyond national ex-post assessment systems (e.g. for hiring, promotions or project selection) where such initiatives can also be taken up, though these lie beyond the scope of this study.

Across the 2010 to 2024 period of our study, it is notable that **no new frameworks for advice-oriented assessment** have been established, save for Australia's EI (which was discontinued after one cycle). From our sample, the overall picture is one of **consolidation of summative-oriented systems** that do not explicitly aim to deliver strategic advice. An exception is the **Norwegian Disciplinary Evaluation**, where since 2011, one of the main tasks of the evaluation committees for biology, medicine and health has been to deliver strategic advice to the Research Council of Norway and the central government. This was also included in the 2017 and 2018 evaluations of social science and humanities, and in the third round of STEM-evaluations, starting in 2021.

Combinations of disciplinary and excellence-oriented criteria are visible in all cases, though how these are combined and play out in different systems varies. Principles of excellence (such as competition through selective performance-based funding based on evaluating traditional research outputs) continue to be emphasised in many systems. Efforts to mitigate or counter perceived dysfunctions in the disciplinary and excellence paradigms have led to visible changes in some but by no means all national assessment systems.

4. Conclusion: pathways & possibilities

A new report from the Global Research Council (May 2025) with examples of how different public funding agencies are advancing RRA

How might elements of the RRA agenda translate differently, according to the methods and rationales of the diverse national assessment systems in our sample? Generally speaking, **principles and criteria of RRA are more readily absorbed in formative, advisory assessment systems**. By contrast, in more competitive summative systems, changes in assessment criteria can directly affect institutional reputation or levels of funding, making changes more consequential, and likely to be contested.

Before the addition of process indicators to systems like the UK REF, evaluation of research inputs and processes was typically the preserve of advisory-oriented assessments. Back in 2009-2015, the **Dutch SEP** included **'vitality and feasibility'** (e.g. staffing, prioritization and project management as performance criteria). By 2021 additional criteria for evaluating included: open science; PhD training; academic culture; and human resources policy.

Incorporating RRA criteria into formative frameworks like the Dutch SEP and Norwegian Disciplinary Evaluation may be less contested, they are also **less likely to act as strong levers for behaviour and culture change**. By contrast **'strong' systems linked to funding seem more likely to shift the priorities of those being assessed,** given the resource and reputational importance of the outcomes (Whitley, 2007). In the UK, the introduction of impact in REF 2014 has contributed to shifts in the status, emphasis and economic rationale for impact-oriented activities across UK higher education, in ways that more discretionary, advisory processes would have struggled to achieve. **Though not without its critics, the inclusion of 'People, Culture and Environment' in REF 2029 will likely generate further strong stimulus for organisations to take these priorities more seriously.**

One uncertainty with respect to RRA-oriented reforms is **what role indicator-based approaches might play** in furthering this agenda. The agreement underpinning CoARA states as its second commitment *"Base research assessment primarily on qualitative evaluation for which peer review is central, supported by responsible use of quantitative indicators."*[3] This has been interpreted by some as an overly-sceptical stance towards the use of bibliometrics, and indicator-based evaluations that use them (e.g. Abramo, 2024).

However, the **potential for indicator-based systems to incorporate elements of the RRA agenda, such as open science, team science, or peer review data, should not be overlooked** (Sivertsen and Rushforth, 2024). Advances in bibliometric meta-data might, for instance, help indicator-based allocation schemes evaluate and monitor certain kinds of open access publishing activities; or track cross-sector or cross-disciplinary collaborations or research impacts. Caution of course needs to be taken in any quantitative assessment system, to ensure data quality, and reliability, and to monitor for unintended consequences.

[3] https://coara.eu/agreement/the-agreement-full-text/

Using advanced bibliometric methods and data to support and advance the RRA paradigm (rather than standing in opposition to it) is an important consideration which merits more analysis and discussion. This is particularly important for countries with established indicator-based national systems, and limited appetite to establish large-scale national peer-review based evaluations.

New data from the **2025 RoRI/Global Research Council (GRC) survey** of GRC-participating funders reinforces the sense of a gradual but meaningful shift in the 'centre of gravity' of these debates towards RRA approaches and agendas (Rushforth et al, 2025). But also of **considerable heterogeneity in the speed and scale of assessment reforms** as they are being applied across different systems, with **no single direction of travel**.

Some systems appear likely to consolidate around large-scale competitive peer-review based systems, while others may explore alternatives (as hinted by Adams et al, 2022). Some policymakers may look to indicator-based systems as an option, while others could look towards advisory-oriented assessments or systems where accreditation is an important purpose. For example, a recent EU-level report by Guillet (2025) calls on research culture and environment to become a focus of the quality assurance of research in higher education institutions conducted by external accreditation bodies.

Interest is also growing in **the potential to incorporate AI tools into data-driven assessment methods**, particularly in light of concerns over the swelling administrative burden of large assessment exercises. The RoRI/GRC 2025 survey of global funders reinforces the openness of many funders to AI tools, albeit tempered by a concern to ensure that these don't introduce fresh biases or problematic incentives (Adams et al., 2022).

Jonathan Adams of Clarivate argues that the UK REF has outlived its usefulness and delivers diminishing returns.

An alternative to the use of national research assessment and funding systems is to rely more on **performance agreements, a 'dialogue-based funding' model,** whereby governments or intermediaries negotiate targets with higher education and research organisations, with funding tied partially or fully to organisations meeting the contractual targets (Salmi and Hauptman, 2006, 40, 58).

Supporters of performance agreements argue they can support diversification of organisational profiling, enhance sensitivity of performance criteria towards institutional missions and public values, generate more context-specific indicators, and strengthen trust between government and institutions

(Jongbloed et al., 2020). **They are not without potential drawbacks though**, with critics highlighting concerns over an erosion of organisational autonomy, increased bureaucracy, and financial penalties for unmet targets occurring for reasons beyond an organisation's control (Sivertsen, 2023).


DAILY NEWSPAPER OF THE YEAR

THE TIMES

SCOTLAND

Elizabeth Day
My toxic dating disaster

Boris Becker
I lost it all but survived prison

Alarm over diversity push

Universities could lose research funding if they fail to prioritise inclusivity

INSIDE TODAY

Alba
Your guide to life in Scotland

A February 2025 front cover story in The Times newspaper in the UK thrust REF reforms into the limelight.

Whichever models are used, **pathways towards RRA reforms in national assessment systems are likely to remain multiple and complex.** Some systems may resist these ideas more actively, for ideological or pragmatic reasons, while others press accelerate more rapidly. Some systems may experience 'pressure from below', in support or opposition to these agendas. Current UK proposals for the 'people, culture and environment' strand of the REF have elicited both, with a heated debate across the university and research community about the merits, pitfalls and practical feasibility of giving these elements greater emphasis in REF 2029.

In 2024, 36% (1147 institutions) of the total population of organisational signatories of DORA were from Latin America; but there are still perceptions in some quarters that the RRA paradigm is European-dominated. Under-resourced institutions in the Global South may also shift back towards the excellence paradigm as a logical response to intense competitive pressures for resources and talent.


US science is under attack: can it survive Trump 2.0?

President Donald Trump and his administration have gutted science agencies, terminated research programmes and cancelled billions of dollars in grants to universities. What are the long-term impacts for the United States and the world?

By Jeff Tollefson, Dan Garisto & Heidi Ledford

Nature on proposed cuts to US federal R&D investment (April 2025)

There is perhaps **a fork in the road over the next 5-10 years**, as some national systems decide to accommodate or further embed an RRA-oriented agenda, while others consolidate along disciplinary/excellence lines. And these developments are of course highly susceptible to larger system shocks (as we are now seeing in the severe cuts proposed to federal R&D budgets in the United States). Any or all of geopolitics, financial crises, austerity, pandemics, and ideological battles over universities may alter the landscape for publicly-funded research in unpredictable ways. The emerging yet still-fragile paradigm of RRA may benefit or be derailed by such perturbations.

Headlines of this paper and further research

This paper has **two objectives:**

- to develop and apply a comprehensive typology of national research assessment and funding systems across thirteen countries (including some from the Global South);
- to analyze changes in these systems from 2010 to 2024. In doing so we also sought to explore potential developments in such systems over the next 5-10 years.

Our typology confirms the **considerable diversity in national assessment and funding systems** reported in existing comparative literature (Geuna and Martin, 2003, Whitley and Gläser, 2007, Hicks, 2012, Zacharewicz et al., 2019, Oschner et al., 2021), reinforcing that there is no universal recipe or template to follow.

Our typology adds several **new contributions**. First, it provides an up-to-date framework for the comparison of diverse systems against **three performance paradigms, four core aspects and four subsidiary aspects**. This makes it easier to compare systems that would otherwise be hard to analyse alongside one another.

Second, our study included countries with at least one of the following types of ex-post national assessment systems currently (or recently) in place: indicator-based funding; peer review linked to funding; peer review linked to organisational improvement; and/or individual-level national assessment systems. By expanding our inclusion criteria to include – for instance – ex-post evaluation systems with the primary assigned purpose of generating strategic advice, we present a **more comprehensive picture of available options for administering funding, measuring performance and assuring quality** across national research systems (c.f. Sivertsen, 2023).

Third, **our sample allows us to shed further light on ex-post systems where the unit of assessment is individual-level research performance**. These predominate in Latin America and have been relatively under-studied in comparative analyses and research evaluation literature more generally (Vasen et al., 2023).

Fourth, **our collaborative model of research and co-authorship** with colleagues based across these thirteen countries has enabled our typology **to draw on and synthesize a far larger set of data, documents and other contextual information** (in languages other than English) that otherwise would not be picked up – for example in a more conventional literature review.

Furthermore, our **longitudinal perspective** has led us to theorize changes in rationales and expectations of national research assessment systems from 2010 to 2024. This is a departure from the cross-sectional designs which typically inform such comparisons. This approach leads us to argue that, over time, **three paradigms of performance assessment** have shaped these systems, to varying degrees.

First, the **disciplinary paradigm** (peer autonomy, internal disciplinary standards); then, the **excellence paradigm** (competition, selectivity of funding, performance indicators) and more recently **the nascent 'responsible research assessment' (RRA) paradigm** (broadening assessment beyond traditional outputs, diversifying research quality, widening public values).

Importantly, we observe that **national systems have not moved seamlessly from one phase to another:** new paradigmatic ideas accumulate over time, creating 'layering' effects, where elements of earlier paradigms persist alongside newer paradigms and combine to generate systems with complex, hybrid characteristics.

Our analysis also shows that t**he RRA agenda is a paradigm still in its early stages**, featuring in some but not all national systems, meaning it is not yet a global trend. **Non-linear and geographically diverse responses to RRA suggest a potential decision-point in the next 5-10 years,** whereby some systems accommodate and adapt to this paradigm in a proactive fashion, while others remain static or move in a different direction. **Even where the RRA paradigm is gaining ground, it remains heterogeneous, in its infancy, and highly context-dependent.**

Lord Vallance, UK Minister for Science & Innovation and Denys Kurbatov, Ukraine's Deputy Minister for Education and Science, at a meeting in July 2025 to draw lessons from the UK REF for a newly-proposed Ukrainian research assessment system.

Our analysis leads us to **caution against oversimplification** when describing trends in national assessment systems. While systems have been paused or discontinued in some countries, others have been introduced or are under consideration for the first time (for example, in Ukraine).

The overall picture is one of **consolidation of such systems under dynamic steady state conditions**, and of **gradual, uneven changes in the rationales, design and objects of assessments**, rather than wholescale transformation. Looking to the near future, the biggest challenge for reforming new and established systems alike will be **selecting appropriate criteria, methods and indicators for robust assessment of research performance** and then **translating aspirational RRA principles into workable assessment methods**. This is the debate now raging in the UK over the people, culture, environment (PCE) strand of REF 2029.

How the emerging RRA paradigm is translated into large-scale, periodic peer review-based assessments — versus advisory assessments; indicator-based systems; individual assessments; or institutional performance agreements – and how the emerging paradigm interacts with existing ideas and practices of disciplinarity and excellence – are **urgent policy questions which demand more empirical and theoretical focus,** both in academic metascience, STS and studies of evaluation, and from reform movements within science, such as CoARA, DORA and the Barcelona Declaration.

Future research should **extend our current analysis to an even broader range of countries not yet included in this study.** Additionally, investigating the practical implementation and outcomes of RRA principles in various national contexts would provide valuable insights for policymakers, institution leaders, and researchers worldwide.

It is essential that **the evolution of these systems reflects each country's unique strategic strengths, needs and priorities** – rather than uncritically pursuing a singular 'best practice' model. In this respect, our typology and analysis of system changes offers opportunities for cross-national mutual learning and further underlines the importance of a comparative research agenda of the kind RoRI has developed through its AGORRA project. Comparative perspectives on these questions are rare, and longitudinal comparative studies even rarer.

This working paper **shines a comparative light on the variegated landscape of national research assessment and funding.** We hope that our new typology will inform more effective and context-appropriate decision-making by those involved in performance evaluation and funding allocation. The ultimate test will be whether these systems continue to sustain, reflect and project the vitality and dynamism of the researchers and research institutions on which so much depends.

5. References

Aagaard, Kaare. 2017. 'The Evolution of a National Research Funding System: Transformative Change through Layering and Displacement'. *Minerva* 55 (3): 279–97.

Abramo, Giovanni. (2024). The forced battle between peer-review and scientometric research assessment: Why the CoARA initiative is unsound, *Research Evaluation*, 2024; rvae021, https://doi.org/10.1093/reseval/rvae021

ACOLA & Office of the Chief Scientist. 2023. Research Assessment in Australia: Evidence for Modernisation. https://www.chiefscientist.gov.au/ResearchAssessment

Adams, Jonathan, Ryan Beardsley, Lutz Bornmann, Jonathan Grant, Martin Szomszor, and Kate Williams. 2022. 'Research Assessment: Origins, Evolution, Outcomes'. Clarivate. https://clarivate.com/academia-government/wp-content/uploads/sites/3/dlm_uploads/XBU968048850-ISI-Research-Assessment-Report-v5b-Spreads.pdf.

Barcelona Declaration. 2024. 'Barcelona Declaration on Open Research Information'. 2024.

Beigel, Fernanda. 2025. 'The Transformative Relation between Publishers and Editors: Research Quality and Academic Autonomy at Stake'. *Quantitative Science Studies*, 1–17.

Benamara, Abdelmajid;, Ahmed; Fahal, Alicia; Kowaltowski, Anh-Khoi; Trinh, Anne; Cody, Catriona; Firth, and et al. 2024. 'Dimensions of Responsible Research Assessment (Full Report and Summary)'. *Figshare: Online Resource*. https://doi.org/10.6084/m9.figshare.26064223.v3.

Bozeman, Barry, and Daniel Sarewitz. 2011. 'Public Value Mapping and Science Policy Evaluation'. *Minerva* 49:1–23.

Braun, Dietmar. 1998. 'The Role of Funding Agencies in the Cognitive Development of Science'. *Research Policy* 27 (8): 807–21.

Braun, Dietmar, and David H. Guston. 2003. 'Principal-Agent Theory and Research Policy: An Introduction'. *Science and Public Policy* 30 (5): 302–8

Bush, Vannevar. 1945. *Science The Endless Frontier: A Report to the President by Vannevar Bush, Director of the Office of Scientific Research and Development, July 1945*. United States Government Printing Office, Washington DC

Butler, Linda. 2007. 'Assessing University Research: A Plea for a Balanced Approach'. *Science and Public Policy* 34 (8): 565–74.

Capano, Giliberto. 2023. 'Ideas and Instruments in Public Research Funding'. In *Handbook of Public Funding of Research*, 73–89. Cheltenham: Edward Elgar Publishing.

Carson, Marcus, Tom R. Burns, and Dolores Calvo. 2009. 'Introduction'. In *Paradigms in Public Policy: Theory and Practice of Paradigm Shifts in the EU*, edited by Marcus Carson, Tom R. Burns, and Dolores Calvo, 11–28. Berlin: Peter Lang.

CLACSO. 2022. 'Declaration of Principles'. Presented at the CLACSO´s XXVII General Assembly, Mexico. https://biblioteca-repositorio.clacso.edu.ar/bitstream/CLACSO/169747/1/Declaration-of-Principes.pdf.

Curry, S., S. De Rijcke, A. Hatch, Dorsamy (Gansen) Pillay, I. van der Weijden, and James Wilsdon. 2020. 'The Changing Role of Funders in Responsible Research Assessment'.

Curry, S., E. Gadd, and James Wilsdon. 2022. 'Harnessing the Metric Tide: Indicators, Infrastructures & Priorities for UK Responsible Research Assessment. Report of The Metric Tide Revisited Panel'.

Dahler-Larsen, P. 2014. Constitutive effects of performance indicators: Getting beyond unintended consequences. Public Management Review, 16(7), 969–986.

Debackere, Koenraad, E. Arnold, G. Sivertsen, J. Spaapen, and D. Sturn. 2018. 'Mutual Learning Exercise: Performance-Based Funding of University Research'. *Directorate-General for Research and Innovation, European Commission, B-1049 Brussels ISBN*, 978–92.

Digital Science. n.d. 'The Nature, Scale and Beneficiaries of Research Impact: An Initial Analysis of Research Excellence Framework (REF) 2014 Impact Case Studies.' http://dera.ioe.ac.uk/22540/1/Analysis_of_REF_impact.pdf.

DORA. 2013. 'The Declaration'. 2013.

Elzinga, Aant. 2012. 'Features of the Current Science Policy Regime: Viewed in Historical Perspective'. *Science and Public Policy* 39 (4): 416–28.

European Commission. 2012. 'A Reinforced European Research Area Partnership for Excellence and Growth'.

European Commission. 2024. 'Action Plan by the Commission to implement the ten commitments of the Agreement on Reforming Research Assessment (ARRA)'.
Expert Review Panel. n.d. 'VQR 2015-2019'. https://www.anvur.it/sites/default/files/2025-02/Expert-Review-Panel_Report-on-VQR-2015-2019.pdf.
Galderisi, Claudio, Mauro Perretti, Nura Galles, and Thed N van Leeuwen. 2019. 'Report of the Group of Experts Charged by ANVUR to Advice on the Process "Valutazione Della Qualità Della Ricerca (VQR)"'. https://www.anvur.it/sites/default/files/2025-02/High-Experts-Report-on-VQR.pdf.
Geuna, Aldo, and Ben R. Martin. 2003. 'University Research Evaluation and Funding: An International Comparison'. *Minerva* 41 (4): 277–304.
Gras, Natalia. n.d. 'Forms of Research Assessment Oriented at Development Problems. Practices and Perspectives from National Science and Technology organisations and Higher Education Institutions in Latin America and the Caribbean.' https://doi.org/10.5281/zenodo.6607850.
Guillet, Sophie. 2025. 'External Quality Assurance of Research in Higher Education Institutions. Taking Stock of the Practices of European Quality Assurance Agencies'. *European Association for Quality Assurance in Higher Education*. https://www.enqa.eu/wp-content/uploads/ENQA-QA-of-R-report.pdf.
Hall, Peter. 1993. 'Policy Paradigms, Social Learning, and the State: The Case of Economic Policymaking in Britain' Comparative Politics, 25 (3): 275-296. http://www.jstor.org/stable/422246
Henriques, Luisa, and Philippe Larédo. 2013. 'Policy-Making in Science Policy: The "OECD Model" Unveiled'. *Research Policy* 42 (3): 801–16.
Hicks, Diana. 2012. 'Performance-Based University Research Funding Systems'. *Research Policy* 41 (2): 251–61.
Hicks, Diana, Paul Wouters, Ludo Waltman, Sarah De Rijcke, and Ismael Rafols. 2015. 'Bibliometrics: The Leiden Manifesto for Research Metrics'. *Nature* 520 (7548): 429–31.
IAP-GYA-ISC. 2023. 'The Future of Research Evaluation'. Paris: Center for Science Futures.
Inglis, Matthew, Elizabeth Gadd, and Elizabeth Stokoe. 2024. 'What Is a High-Quality Research Environment? Evidence from the UK's Research Excellence Framework'. *Research Evaluation*. https://doi.org/10.1093/reseval/rvae010.
Jones, Richard and Wilsdon, James. (2018). *The biomedical bubble: Why UK research and innovation needs a greater diversity of priorities, politics, places and people*. London, Nesta.
Jong, Lisette; Franssen, Thomas; Pinfield, Stephen (2021). 'Excellence' in the Research Ecosystem: A Literature Review. (RoRI Working Paper No. 5). Research on Research Institute. Report. https://doi.org/10.6084/m9.figshare.16669834.v2
Jongbloed, Ben, and Harry de Boer. 2020. 'Performance Agreements in Denmark, Ontario and the Netherlands'.
Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. pp. 54.University of Chicago Press: Chicago
Lepori, Benedetto, Peter van den Besselaar, Michael Dinges, Bianca Potì, Emanuela Reale, Stig Slipersæter, Jean Thèves, and Barend van der Meulen. 2007. 'Comparing the Evolution of National Research Policies: What Patterns of Change?' *Science and Public Policy* 34 (6): 372–88. https://doi.org/10.3152/030234207X234578.
Leslie, Ian (2025). 'Notes on the Great Vibe Shift'. Substack, 18 January 2025. https://www.ian-leslie.com/p/notes-on-the-great-vibe-shift
Martin, Ben R. 2011. 'The Research Excellence Framework and the "Impact Agenda": Are We Creating a Frankenstein Monster?' *Research Evaluation* 20 (3): 247–54.
Ochsner, Michael, Emanuel Kulczycki, Aldis Gedutis, and Ginevra Peruginelli. 2020. 'National Research Evaluation Systems'. In *Handbook Bibliometrics*, edited by Rafael Ball, 99–106. De Gruyter Saur. https://doi.org/10.1515/9783110646610-011.
REF. (2029). Research Excellence Framework. https://2029.ref.ac.uk/?detectflash=false
Rijcke, Sarah de, Paul Wouters, Alex Rushforth, Thomas Franssen, and Björn Hammarfelt. 2016. 'Evaluation Practices and Effects of Indicator Use—a Literature Review'. *Research Evaluation* 25 (2): 161–69.
Rip, Arie. 2000. 'Fashions, Lock-Ins and the Heterogeneity of Knowledge Production'. In *The Future of Knowledge Production in the Academy*, 28–39. Society for Research into Higher Education and Open Universi.

———. 2004. ‘Strategic Research, Post-Modern Universities and Research Training’. *Higher Education Policy* 17:153–66.
Rushforth, Alexander, and Björn Hammarfelt. 2023. ‘The Rise of Responsible Metrics as a Professional Reform Movement: A Collective Action Frames Account’. *Quantitative Science Studies* 4 (4): 879–97. https://doi.org/10.1162/qss_a_00280.
Salmi, Jamil, and Arthur Hauptman. 2006. ‘Innovations in Tertiary Education Financing: A Comparative Evaluation of Allocation Mechanisms’. *World Bank*, Education Working Paper Series Number 4.
Scholten, Wout, Thomas P. Franssen, Leonie van Drooge, Sarah de Rijcke, and Laurens K. Hessels. 2021. ‘Funding for Few, Anticipation among All: Effects of Excellence Funding on Academic Research Groups’. *Science and Public Policy* 48 (2): 265–75.
Sivertsen, Gunnar. 2018. ‘The Norwegian Model in Norway’. *Journal of Data and Information Science* 3 (4): 3–19.
———. 2023. ‘Performance-Based Research Funding and Its Impacts on Research organisations’. In *Handbook of Public Funding of Research*, 90–106. Edward Elgar Publishing.
Sivertsen, Gunnar, and Alexander Rushforth. 2024. ‘The Ongoing Reform of Research Assessment’. In *Challenges in Research Policy*, edited by Gunnar Sivertsen and Liv Langfeldt. Dordrecht: Springer.
Stern, Lord Nicholas. 2016. ‘Building on Success and Learning from Experience. An Independent Review of the Research Excellence Framework’. London: Department for Business, Energy and Industrial Strategy. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/541338/ind-16-9-ref-stern-review.pdf.
Thomas, Duncan A., Maria Nedeva, Mayra M. Tirado, and Merle Jacob. 2020. ‘Changing Research on Research Evaluation: A Critical Literature Review to Revisit the Agenda’. *Research Evaluation* 29 (3): 275–88.
UNESCO. 2021. ‘Recommendation on Open Science’.
Vasen, Federico, Nerina F. Sarthou, Silvina A. Romano, Brenda D. Gutiérrez, and Manuel Pintos. 2023. ‘Turning Academics into Researchers: The Development of National Researcher Categorization Systems in Latin America’. *Research Evaluation* 32 (2): 244–55.
Wagner, Caroline S, Irene Brahmakulam, Brian Jackson, Anny Wong, and Tatsuro Yoda. 2001. ‘Science and Technology Collaboration: Building Capacity in Developing Countries: Prepared for the World Bank’. *RAND Corporation*.
Whitley, Richard. 2000. *The Intellectual and Social organisation of the Sciences*. Oxford: OUP.
———. 2007. ‘Changing Governance of the Public Sciences’. In *The Changing Governance of the Sciences*, 3–27. Dordrecht: Springer.
———. 2019. ‘Changing Science Policies, Authority Relationships and Innovations in Public Science Systems’. In *Handbook on Science and Public Policy*, 204–26. Edward Elgar Publishing.
Whitley, Richard, and Jochen Gläser. 2007. *The Changing Governance of the Sciences*. Dordrecht: Springer.
Williams, Kate, and Jonathan Grant. 2018. ‘A Comparative Review of How the Policy and Procedures to Assess Research Impact Evolved in Australia and the UK’. *Research Evaluation* 27 (2): 93–105. https://doi.org/10.1093/reseval/rvx042.
Wilsdon, James. 2016. *The Metric Tide: Independent Review of the Role of Metrics in Research Assessment and Management*. HEFCE.
Wróblewska, Marta Natalia. 2025. ‘One size fits all? A comparative review of policy-making in the area of research impact evaluation in the UK, Poland and Norway’. *Research Evaluation*, Volume 34, rvaf010, https://doi.org/10.1093/reseval/rvaf010.
Xiaoxuan, L., & Fang, X. U. (2020). How to Break the ‘Siwei’?—Practice and Enlightenment Based on Research Institute Evaluation of Chinese Academy of Sciences. *Bulletin of Chinese Academy of Sciences (Chinese Version)*, *35*(12), 1431–1438. https://doi.org/10.16418/j.issn.1000-3045.20201116002
Zacharewicz, Thomas, Benedetto Lepori, Emanuela Reale, and Koen Jonkers. 2019. ‘Performance-Based Research Funding in EU Member States—a Comparative Assessment’. *Science and Public Policy* 46 (1): 105–15.

Zhang, L., & Sivertsen, G. (2023). The New Research Assessment Reform in China and Its Implementation. In Towards a New Research Era (pp. 239–252). Brill. https://doi.org/10.1163/9789004546035_017
Ziman, John M. 1994. *Prometheus Bound*. Cambridge: Cambridge University Press.

Appendix

National Research Assessment and Funding Systems

Country	Name of system(s)	Year of introduction / major change(s)	Census period	Summary
Argentina	Program for Researcher Teachers (PRINUAR, previously PROINCE)	1994	2 yrs (admission and promotion) / 4 years (permanence)	PRINUAR - and its predecessor PROINCE – evaluates individual researchers. It was created to promote research activities in universities by categorizing faculty members based on their research output. The process follows informed peer review and aims to incentivize performance through categorizing researchers (reputation) and providing small amounts of funding based on performance.
	Institutional Evaluation Program /Programa de Evaluación Institucional (PEI)	2005	Ad hoc	Voluntary program introduced to help institutions assess their R&D activities and produce strategic advice to improve their performance (formative assessment). Resources allocated on the basis of PEI results are limited. PEI follows informed peer review (use of bibliometrics varies per discipline). Evaluations also include a self-assessment and site visit by an external evaluation committee.
	CONICET Career of Scientific and Technological Researcher (CICYT)	1961	1 or 2 years according to seniority	Individual career system whose members are constantly evaluated for performance. Performance report outcomes are generally consequential for individuals' reputations affecting career advancement prospects in CONICET program. Evaluations use peer review informed in some disciplines by bibliometrics and other types of quantitative information.
Australia	Excellence in Research for Australia (ERA)	2010, Ceased since 2023	3 years	ERA identified and promoted excellence across the full range of research in Australia's higher education institutions. ERA assessed research outputs using a combination of peer review moderated by a panel of experts and bibliometric analysis moderated by a panel of experts to determine results, depending on the field of research being assessed (STEM disciplines predominantly assessed by bibliometric analysis, and SSH disciplines by peer review). ERA impacted institutional reputation and funding, however it ceased being connected to funding in its 2018 exercise and ceased operating at all in 2023.
	Engagement and Impact (EI) Assessment	2018	Not repeated	EI assessed how well researchers engage with end-users of research. It also assessed how well universities are translating research into impacts beyond academia. EI assessed engagement and impact using expert review of case studies and other evidence (including engagement narratives and other indicators). The exercise provided strategic advice to be used locally at each institution.

Brazil	CAPES Evaluation System for Graduate Programs	1976	4 years	The primary means of national research evaluation in Brazil occurs via regular assessments of all graduate programs, carried out by the CAPES federal agency. The stated aims include quality assurance, overviews of educational and research activities, and strengthening the basis of science, technology, and innovation in Brazil. This evaluation process occurs every four years. CAPES performs holistic evaluations, focusing on quality of research activities but also educational performance, such as graduated students employment rates. Evaluation involves informed peer review, conducted by members of the Evaluation Areas - which group graduate programs according to their disciplines. CAPES assigns a quality label to the assessed program (on a scale of 3 to 7, with 3 being the minimum required for a program to operate and 7 representing an internationally competitive program), and provides a small amount of financial support to postgraduate programs: those rated 6 or 7 receive slightly more than those below. Although only a limited resource earmarked for activities such as mobility, conference participation, and publications, results are important for enhancing many programs' research performing capacities.
Chile	National Accreditation Commission / Comisión Nacional de Acreditación (CNA)	2006	3-7 years	The National Accreditation Commission (CNA) evaluates and accredits the quality of higher education institutions, universities, autonomous technical training centres in Chile, and programs they all offer. One of the optional areas of the institutional accreditation process for which it is responsible is research (and creation and innovation). Although a voluntary dimension, performing well on research is essential to obtain the maximum period of accreditation, with the aim of encouraging development and strengthening the quality of research in Chile. The evaluation process consists of internal self-evaluation, carried out by the respective institutions; external evaluation, carried out by peer evaluators; and final judgment issued on the basis of the weighting of the information gathered. Research evaluation-based accreditation by CNA provides the institution with quality support, prestige and recognition, although it is not a condition for an institution to develop research.
China	National Disciplinary Evaluation	2002	4 years	An indicator-informed system that assesses the performance of over 1,100 universities and 31,000 disciplines nationwide. Although its results do not directly determine funding, they significantly influence funding decisions by various stakeholders and internal resource allocation within universities.
	Double First-Class Evaluation	2017	5 years	"Double" refers to both university-level and disciplinary-level assessments. Currently, 147 universities (and more than 300 disciplines) are included in this evaluation, representing about 5% of all higher education institutions and showcasing the top-tier universities in China. The evaluation is performed by committees informed by indicators, with results affecting reputation and funding.
	National assessment and selection systems for elite individual researchers	1994	1 year	Highly competitive selections of elite individual researcher accompanied by substantial resources, including high salaries, generous research funding, and other institutional support. Performed by committees informed by indicators, with results highly consequential for reputation, funding and career advancement prospects of individuals. Systems within the program include the Distinguished Youth Scientist Fund and Yangtze River award.

	Chinese Academy of Science research institute evaluation	1998	1 or 5 years	The research institute evaluation, initiated in 1998, is an important tool for the Chinese Academy of Sciences (CAS) to manage over 100 research institutes, mainly serving to provide guidance, motivation, and measurement. The strategic goals of CAS vary in different development periods, and accordingly, the methods and applications of research institute evaluation also differ. Overall, the evaluation has gone through developmental stages from quantitative indicators to qualitative assessment and then to comprehensive evaluation. The evaluation results are generally linked to the allocation of resources for the institutes (with varying degrees of intensity) and have a significant impact on their reputation.
Colombia	Research Groups Assessment Model / Modelo de Medición de Grupos	1991	~2 years	An assessment of research groups, carried out each 2-3 years by Minciencias. The assessment establishes a system of categorization of research groups, based on their scientific and educational production and societal impact of research. Assessment is based on the quantity of outputs and quality of contributions by research groups and establishes a rank of groups for each knowledge field. The ascribed aim is to gather information about research in Colombia, and to establish competition and quality standards. The categories awarded by Minciencias have an impact on the prestige of universities and supports comparison among them. The model also performs a secondary function of assessing and classifying individual researchers' performance.
	Decreto 1279 de 2002	2002	1 year	Decreto 1279 de 2002 is the main incentive for research in public universities. For each academic paper (and other type of outputs) in journals assessed by Publindex, individual researchers receive several points. Each point has a value in monetary terms, and it increases the monthly salary of the researcher.
	High Quality Accreditation Model	1992	4-10 years	CNA (National Accreditation Council) is an institution that certifies universities and academic programs for their performance. Its national system of accreditation is based on peer review (assessment by experts) of a self-evaluation and takes into account research metrics. The accreditation granted can have a duration of 4, 6, 8, and 10 years.
India	National Institutional Ranking Framework (NIRF)	2016	1 year	An indicator-based system primarily intended to allocate reputation to all public and private colleges and universities in the country. University performance is assessed through a panel of indicators, such as external project revenue, quantitative bibliometrics, number of papers, citations and number of patents. and input counts (e.g., student numbers, staff gender composition). Results have some indirect effects for individuals, departments or universities, influencing promotion and career progression prospects for individual researchers, and use by students to inform choice of educational enrolment; results also have some direct effects in determining eligibility of universities for participating in some competitive funding schemes.
Italy	Evaluation of Research Quality (VQR)	2011	5 years	Combined indicator and peer review based system evaluating the research outcomes of public universities and research institutes, as well as those of private institutions that voluntarily submit their research outcomes for evaluation. Consequential for funding and reputation.

Mexico	National System of Researchers (SNI)	1984	Varies per seniority level	The National System of Researchers (SNI) is a national career system that categorizes individual awardees. Evaluations of persons categorized in the SNI are organised according to areas of knowledge, and carried out through peer review and quantitative indicators linked to publication and other activities. Classification results are consequential for individual reputations and career prospects within SNI and beyond the formal confines of SNI can influence the ability to access resources for research.
Netherlands	Strategy Evaluation Protocol (SEP), previously Standard Evaluation Protocol	1994	6 years	All publicly-funded research performing organisations in the Netherlands are evaluated on criteria, guided by the Strategy Evaluation Protocol (SEP) document. SEP aims to provide organisations with strategic advice via an evaluation report that can inform organisational changes and learning. Results of evaluations are not directly connected to funding outcomes. Evaluation is largely peer review-based, consisting of an initial self-evaluation report by the organisation, submitted to an external expert panel prior to their site visit. Reports from site visits are often made publicly available.
Norway	Evaluations of specific subjects and thematic areas	1990	10 years	Subject specific and thematic area evaluations are conducted every ten years, per subject area. The exercise uses peer review by international expert panels, providing critical reviews of Norwegian research in international perspective and produces strategic advice for increased quality and efficiency. The exercise does not affect institutional funding directly. Reports are published, with basic data and methods made publicly available.
	Indicator-Based Funding	2004, due to cease for universities in 2025	1 year	Indicator Based System allocates a small share of funding for research, according to four indicators of research performance. The system combines research and educational indicators, making no special allocation for research activities. System not running for universities anymore, only hospitals and public research institutes.
Poland	Evaluation of Quality of Scientific Activity / Ewaluacja Jakości Działalności Naukowej (EJDN)	Early 1990s, current format run in 2020/1	4 years	Centralized, periodic evaluation of research activities. Conducted approximately every 4 years by the Ministry for Higher Education. Alongside a points-based system, a peer review-based element was introduced in 2020 to evaluate societal impact, modelled on the innovation to the UK REF 2014. The evaluation informs core funding and affects certain privileges of the institutions (such as granting PhD titles) for the following 4-year long period.
	Algorithm Performance Based Funding System	Early 1990s Current format 2020	1 year	An indicator-based funding system for Polish HE and research institutions. Funding is distributed to organisations based on an algorithm which takes into account several factors, including number of students and tier of institution in the previous EJDN evaluation round (see above).

	Research University Program (IDUB)	2019	6 years	The Research University (IDUB) program, launched by the Polish government in 2019, aims to improve the research performance of top Polish universities. 10 universities were selected from 20 eligible institutions in its first year, receiving a 10% subsidy increase for 2020-2026 to support their development plans. These universities must undergo midterm and final evaluations by an international expert panel, based on informed peer review, with results affecting reputation and funding.
UK	Research Excellence Framework (REF), previously Research Selectivity Exercise (RSE) and Research Assessment Exercise (RAE).	1986 (RSE), 1992 (RAE), 2014 (REF)	~7-8 years	The Research Excellence Framework (REF) is a comprehensive process of expert review that assesses the performance of academic research across 34 subject-specific units of assessment in the UK. Every university department or group can submit to a unit of assessment, where their research is evaluated and scored by specialized panels primarily via peer review. The results of REF panel assessments are used by each of the four devolved nations in the UK to inform the distribution of formulaic block-grant funding, known as 'Quality-related' (QR) funding, to higher education institutions.

researchonresearch.org/