



## METHOD ARTICLE

**REVISED** The SCOPE framework – implementing ideals of responsible research assessment [version 2; peer review: 2 approved]Laura Himanen <sup>1,2</sup>, Erica Conte <sup>3</sup>, Marianne Gauffriau <sup>4</sup>, Tanja Strøm<sup>5</sup>, Baron Wolf <sup>6</sup>, Elizabeth Gadd <sup>7</sup><sup>1</sup>Faculty of Management and Business, Tampere University, Kanslerinrinne 1, FI-33014, Tampere, Finland<sup>2</sup>CSC – IT Center for Science, Keilaranta 14, Espoo, 02101, Finland<sup>3</sup>Unity Health Toronto, 30 Bond St, Toronto, Ontario, L1Z 1P3, Canada<sup>4</sup>IT University of Copenhagen, Rued Langaards vej 7, Copenhagen, DK-2300, Denmark<sup>5</sup>Oslo Metropolitan University, (OsloMet), Pilestredet 46, Oslo, 0167, Norway<sup>6</sup>University of Kentucky, 311 Main Building, Lexington, Kentucky, 40502, USA<sup>7</sup>Loughborough University, Epinal Way, Loughborough, England, LE11 3TU, UK**V2** First published: 28 Sep 2023, 12:1241  
<https://doi.org/10.12688/f1000research.140810.1>Latest published: 17 May 2024, 12:1241  
<https://doi.org/10.12688/f1000research.140810.2>**Abstract****Background**

Research and researchers are heavily evaluated, and over the past decade it has become widely acknowledged that the consequences of evaluating the research enterprise and particularly individual researchers are considerable. This has resulted in the publishing of several guidelines and principles to support moving towards more responsible research assessment (RRA). To ensure that research evaluation is meaningful, responsible, and effective the International Network of Research Management Societies (INORMS) Research Evaluation Group created the SCOPE framework enabling evaluators to deliver on existing principles of RRA. SCOPE bridges the gap between principles and their implementation by providing a structured five-stage framework by which evaluations can be designed and implemented, as well as evaluated.

**Methods**

SCOPE is a step-by-step process designed to help plan, design, and conduct research evaluations as well as check effectiveness of existing evaluations. In this article, four case studies are presented to show

**Open Peer Review****Approval Status**

	1	2
<b>version 2</b>		
(revision)		
17 May 2024		
<b>version 1</b>		
28 Sep 2023		

1. **Sven E. Hug** , University of Zurich, Zürich, Switzerland2. **Dr Maria Alejandra Pinero de Plaza, PhD** , Flinders University, Adelaide, Australia

Any reports and responses or comments on the article can be found at the end of the article.

how SCOPE has been used in practice to provide value-based research evaluation.

## Results

This article situates SCOPE within the international work towards more meaningful and robust research evaluation practices and shows through the four case studies how it can be used by different organisations to develop evaluations at different levels of granularity and in different settings.

## Conclusions

The article demonstrates that the SCOPE framework is rooted firmly in the existing literature. In addition, it is argued that it does not simply translate existing principles of RRA into practice, but provides additional considerations not always addressed in existing RRA principles and practices thus playing a specific role in the delivery of RRA. Furthermore, the use cases show the value of SCOPE across a range of settings, including different institutional types, sizes, and missions.

## Keywords

Research evaluation, responsible research assessment, evaluation framework, SCOPE



This article is included in the [Research on Research, Policy & Culture](#) gateway.



This article is included in the [Innovations in Research Assessment](#) collection.

**Corresponding author:** Elizabeth Gadd ([e.a.gadd@lboro.ac.uk](mailto:e.a.gadd@lboro.ac.uk))

**Author roles:** **Himanen L:** Conceptualization, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Conte E:** Conceptualization, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Gauffriau M:** Conceptualization, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Strøm T:** Conceptualization, Investigation, Methodology, Writing – Review & Editing; **Wolf B:** Conceptualization, Investigation, Methodology, Writing – Review & Editing; **Gadd E:** Conceptualization, Investigation, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** The author(s) declared that no grants were involved in supporting this work.

**Copyright:** © 2024 Himanen L *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Himanen L, Conte E, Gauffriau M *et al.* **The SCOPE framework – implementing ideals of responsible research assessment [version 2; peer review: 2 approved]** F1000Research 2024, 12:1241 <https://doi.org/10.12688/f1000research.140810.2>

**First published:** 28 Sep 2023, 12:1241 <https://doi.org/10.12688/f1000research.140810.1>

**REVISED Amendments from Version 1**

This revision responds to reviewers' comments by revising the title to better describe the scope of the article; by expanding section 2 to make the challenges of defining responsible research assessment clearer and situating the movement within the broader evaluation literature; and by expanding the section describing some of the limitations of the SCOPE framework.

**Any further responses from the reviewers can be found at the end of the article**

**1. Introduction**

This article introduces the SCOPE framework for responsible research assessment (RRA) developed by the International Network of Research Management Societies' (INORMS) Research Evaluation Group (REG) and situates it within the sector-wide drive towards more meaningful and robust research evaluation practices. It begins with a short overview of RRA and some of the underpinning declarations and principles that form its foundation. The need for the SCOPE framework within the research enterprise, as a mechanism which both delivers, and expands upon, those principles is then outlined. The framework is then described and justified with reference to the scholarly literature. Four use cases demonstrating the wide applicability of SCOPE are then provided and some conclusions regarding both its usefulness and its applicability are drawn, highlighting both its strengths and weaknesses.

**2. Moving towards responsible research assessment**

Research and researchers are heavily evaluated. There are many reasons for this. First and foremost, evaluation is a distinctive characteristic of science itself, as the demand for verifiability of results puts scientific research under the exacting scrutiny of fellow experts. As Robert Merton puts it (Merton, 1973, p. 276), "the activities of scientists are subject to rigorous policing, to a degree perhaps unparalleled in any other field of activity". Evaluation has a major role in the knowledge-making process as both a gatekeeper and a legitimator of knowledge, and as a process itself: evaluation sets standards of research quality (Lamont & Huutoniemi, 2011). Demands for greater accountability, linked with diminishing and more competitive funding opportunities, have also created expectations for universities to be both efficient and accountable. As such evaluation is essential to ensure that funding is used in the best possible way (Geuna & Martin, 2003). Research evaluation can also be seen as a governance tool to improve the quality of scholarship (Gläser & Laudel, 2007). It is used for monitoring the activities of researchers, projects, programmes, departments, and institutions – even nations.

According to Geuna & Martin research evaluations tend to focus on four typical output measures: volume, quality, impact, and utility, with peer review and bibliometric measures as their main methods (Geuna & Martin, 2003). While peer review is considered as the key evaluation mechanism of scholarly work (Birukou *et al.*, 2011), over the last two decades we have witnessed a strong increase in the use of quantitative techniques (Van Leeuwen, 2004). As stated in the Leiden Manifesto (Hicks *et al.*, 2015, p. 431), "research metrics can provide crucial information that would be difficult to gather or understand by means of individual expertise." Thelwall *et al.* argue that certain dimensions of scientific quality, such as scientific and societal impact and visibility, can be well-served by indicators (Thelwall *et al.*, 2015). However, others have argued that indicators used in researcher assessment do not capture the elements that reflect high-quality research and knowledge advancement (Aubert Bonn & Pinxten, 2021b).

Whatever method is used, it is apparent that the consequences, both intended and unintended, of evaluating the research enterprise and particularly individual researchers are considerable. Some of the negative consequences include, but are not limited to, the stress and burden on researchers and administrators, tension between colleagues, biased and unfair evaluations, and misalignment with organisational missions and values (see, e.g., Aubert Bonn *et al.*, 2022; Aubert Bonn & Pinxten, 2021a; Benedictus *et al.*, 2016; Gadd, 2021; Lebel & McLean, 2018; Moher *et al.*, 2018; Morrish, 2017; Muller, 2018; Saenen *et al.*, 2019; Wellcome Trust, 2020).

As pointed out by Rushforth & Hammarfelt, within the field of quantitative science studies, as well as research policy context, the concerns over the limitations of evaluative bibliometrics are almost as old as the tools (Rushforth & Hammarfelt, 2023). But it has taken some decades for these concerns to become more widely acknowledged. Curry *et al.* consider the wider spread of critical reflections on research assessment as a response to an intensified debate around how research is funded and practiced, and how research cultures could be made more inclusive and impactful. They go on to point out that while public and private research funding continues to grow, the investment now comes with heightened expectations for mission-oriented research, open scholarship, interdisciplinarity, diversity and inclusion (Curry, Stephen *et al.*, 2020). All of these issues have been found to be worsened by largely bibliometric research assessment approaches. In response, over the last decade several guidelines and principles have been published to support moving towards more responsible modes of research evaluation, now commonly referred to as RRA.

## 2.1 Key RRA principles and guidelines

An often adopted definition of RRA proposed by Curry et al., is that it “is an umbrella term for approaches to assessment which incentivise, reflect and reward the plural characteristics of high-quality research, in support of diverse and inclusive research cultures” (Curry, Stephen et al., 2020, p. 4). But, as pointed out by Gauffriau, so far no exact unifying definition exists (Gauffriau, 2023), so this section presents four internationally well-known examples of principles and guidelines that seek to codify RRA for stakeholder communities.

In 2012, during the Annual Meeting of The American Society for Cell Biology in San Francisco, a group of editors and publishers of scholarly journals developed a set of recommendations known as the San Francisco Declaration on Research Assessment (DORA) (DORA, 2012). Since then, it has become a worldwide initiative, with over 23,000 individual signers and close to 2,900 organizational signers as of June 2023. DORA consists of 18 recommendations that are aimed at funding agencies, academic institutions, journals, organizations that supply metrics, and individual researchers. Even though it highlights the need to consider the value and impact of all research outputs in assessment, the recommendations focus primarily on practices relating to peer reviewed journal articles as the central mechanism by which research is currently assessed. A major recommendation is to eliminate the use of journal-based metrics and to assess research on its own merits rather than on the basis of the journal in which it is published. It also recommends the need to use a broad range of impact measures including qualitative indicators. For publishers and researchers, the recommendations also encompass responsible authorship practices, and for publishers and organizations that supply metrics related to issues of openness and transparency be included.

In 2015, the Leiden Manifesto for research metrics was published providing ten principles to guide bibliometric-based evaluation (Hicks et al., 2015). The authors of the manifesto were alarmed by the pervasive misapplication of indicators to the evaluation of scientific performance, and as a response presented the ten principles, “so that researchers can hold evaluators to account, and evaluators can hold their indicators to account” (ibid., 430). The first three principles consider the role of metrics in research evaluation on a general level. They remind us that the role of quantitative metrics in research assessment should be to support qualitative, expert assessment. Indicators should not be allowed to substitute informed judgement. When indicators are used, they should consider diverse research missions, and measure performance against those missions. Also, using metrics in research assessment poses a risk to locally relevant research as in many parts of the world, research excellence is equated with English-language publications. The remaining seven principles are more practical in nature, considering issues around data collection and analytical processes, the use of indicators and the effects they have on the system.

Shortly after the Leiden Manifesto was produced, an independent review of the role of metrics in research assessment and management was published called The Metric Tide (Wilsdon et al., 2015). The review examined the role of metrics in the UK Research Excellence Framework (REF) conducted in 2014, but it also explored wider issues by looking at the applicability of metrics within different research systems, comparing the peer review system with metric-based alternatives, as well as examining the effects of the growing use of quantitative indicators on different aspects of research culture.

As part of the review, building on the concept of ‘responsible research and innovation’ the authors propose the notion of ‘responsible metrics’ as a way of framing appropriate uses of quantitative indicators in, inter alia, the assessment of research. Their understanding of responsible metrics is built on five principles that have some commonality with elements of DORA and the Leiden Manifesto. They call for recognition that quantitative evaluation should support expert assessment, for basing metrics on the best possible data in terms of accuracy and scope and keeping data collection and analytical processes open and transparent. Accounting for variation by field and using a range of indicators is recommended, as well as recognizing and anticipating the systemic and potential effects of indicators.

The Hong Kong Principles for assessing researchers: Fostering research integrity was published in 2020 (Moher et al., 2020). The starting point for the Hong Kong Principles differed from the three sets of principles previously described, in that its focus is on the need to recognize and reward researchers for behaviours that strengthen research integrity. The authors state that for knowledge to benefit research and society, it must be trustworthy, robust, rigorous, and transparent. The five principles call for researchers to be assessed on accurate and transparent research reporting and engaging with open science practices. In line with DORA, the Leiden Manifesto and the Metric Tide, the Hong Kong Principles call for valuing a broad range of research and scholarship, such as replication, innovation and translation, and other contributions to responsible research and scholarly activity, such as peer review activity, mentoring and outreach.

The most recent addition to the responsible research assessment landscape is the Agreement on Reforming Research Assessment which encompasses many of the ambitions of earlier declarations and expands upon them by requiring

institutions to commit to actually changing their practice within an agreed timeframe (European University Association *et al.*, 2022). The Agreement was published in 2022, and more than 350 organisations from over 40 countries were involved in the drafting. Signing up to the agreement became possible in September 2022 and by July 2023 almost 600 organisations had signed. The Agreement sets a shared direction for changes in assessment practices, as well as a timeframe for implementing reforms. In signing, organisations make four core commitments: to recognize a broader diversity of outputs, practices and activities when assessing research; to base assessment primarily on qualitative judgement supported by quantitative indicators where appropriate; to avoid inappropriate uses of journal and publication metrics and to avoid using university rankings in researcher assessment.

Much of the defining of RRA takes place through declarations and principles, which are based mainly on the experiences of expert practitioners, but moving forward the field of RRA might benefit from more well-defined and tested methods from other research fields such as for example but not limited to Program Evaluation (see, e.g. Rossi *et al.*, 2019) and Utilization-Focused Evaluation (see, e.g. Patton & Campbell-Patton, 2022) from evaluation science, and Fast-and-Frugal Heuristics (see, e.g. Bornmann & Marewski, 2019) from decision science.

### 3. The SCOPE Framework and RRA principles

In 2001 the International Network of Research Management Societies’ (INORMS) was formed to bring together research management societies and associations from across the globe. In recognition of the fact that research assessment was having a growing influence on the research management profession, INORMS established a Research Evaluation Group (INORMS REG) in 2018 to consider how best to ensure that research evaluation is meaningful, responsible, and effective. As part of the INORMS REG’s aim of guiding university leaders and practitioners in the adoption and practice of responsible research evaluation, they developed a framework that both enabled evaluators to deliver on existing principles of responsible assessment and to address some additional critical elements. As such, the SCOPE framework is a practical, five-stage step-by-step process for evaluating responsibly, supported by three overarching principles. Table 1 outlines how SCOPE seeks to deliver on some of the key elements of existing initiatives.

SCOPE bridges the gap between principles and their implementation by providing a structured and orderly framework by which evaluations can be designed and implemented as well as evaluated. Existing principles focus mainly on either evaluating a specific entity, like researchers in the case of DORA and Hong Kong Principles, or via a particular mechanism, like research metrics in the case of Leiden Manifesto and Metric Tide. SCOPE seeks to be applicable across the whole research ecosystem, enabling a responsible approach to evaluating any entity via any relevant mechanism.

As well as enabling the implementation of existing RRA principles, SCOPE also brings additional elements perceived to be necessary in the implementation of responsible assessments. These include three essential principles: 1) to evaluate with the evaluated; 2) to evaluate only where necessary; and 3) to evaluate with the appropriate expertise. It also addresses the need to be context-sensitive, to consider both qualitative and quantitative options, and to probe for unintended

**Table 1. The relationship between SCOPE and RRA principles.**

Principle/Declaration	Objectives	Relationship with SCOPE
<p><b>San Francisco Declaration on Research Assessment (DORA), 2012</b> Seeks improved assessment of researchers and a better scholarly communication ecosystem.</p>	<p>18 recommendations for different stakeholders. The key themes are to:</p> <ul style="list-style-type: none"> <li>• Eliminate journal-based metrics.</li> <li>• Assess research on its own merits.</li> <li>• Take advantage of online publication possibilities.</li> </ul>	<p>SCOPE shares DORA’s vision for better researcher assessment &amp; eliminating the poor use of journal metrics, but is broader in focus, overseeing the responsible assessment of any entity.</p>
<p><b>Leiden Manifesto, 2015</b> Seeks more responsible use of bibliometrics in research assessment.</p>	<p>10 principles for the responsible use of bibliometrics in research assessment focused on:</p> <ul style="list-style-type: none"> <li>• Metrics supporting rather than supplanting expert assessment.</li> <li>• Mission-based performance assessment.</li> <li>• Accounting for variation by field in citation metrics.</li> </ul>	<p>SCOPE shares the Leiden Manifesto’s vision for contextualized use of bibliometrics, but is not limited to quantitative indicators, as it accounts for qualitative measures too.</p>

**Table 1.** Continued

Principle/Declaration	Objectives	Relationship with SCOPE
<p><b>The Metric Tide, 2015</b> Seeks to guide a broad range of research assessment approaches.</p>	<p>Five principles for all forms of research assessment:</p> <ul style="list-style-type: none"> <li>• Robustness</li> <li>• Humility</li> <li>• Transparency</li> <li>• Diversity</li> <li>• Reflexivity</li> </ul>	<p>SCOPE also has a broad focus, but does not stop at providing principles, as it also provides a pragmatic, step-by-step process for evaluating responsibly that includes characteristics like value-led beginnings and a sense-checking probe stage.</p>
<p><b>Hong Kong Principles, 2020</b> Seeks to reward practices that lead to researcher integrity rather than unhelpful &amp; limited publication-based rewards.</p>	<p>Series of principles for assessing researchers that reward research integrity focused on:</p> <ol style="list-style-type: none"> <li>1. Assessing responsible research practices</li> <li>2. Valuing complete reporting</li> <li>3. Rewarding the practice of open science</li> <li>4. Acknowledging a broad range of research activities</li> </ol>	<p>SCOPE also offers value-based assessments, but does not prescribe what those values should be, instead letting the evaluators (together with the evaluated) generate the values that are most meaningful to them.</p>

consequences and to evaluate your evaluation. A full [19-page guide to SCOPE](#) is available on the INORMS REG webpage ([International Network Of Research Management Societies-Research Evaluation Group, 2021](#)) and it is not our purpose to reproduce that here. However, in the next section we provide an outline of the SCOPE framework and principles in some detail situating it within the existing literature.

#### 4. The principles of SCOPE

##### 4.1 Evaluate only where necessary

The five stages of SCOPE presented in [Figure 1](#) operate under three main principles. The first is to evaluate only where necessary. Hallonsten argues that science has been enormously productive even in times when quantitative performance



**Figure 1.** The SCOPE Framework (from [The SCOPE Framework: a five-stage process for evaluating responsibly: <https://doi.org/10.26188/21919527.v1>](#)).

evaluation was not a tool for science policy or university governance, that is, for most of modernity (Hallonsten, 2021). He goes on to conclude that whilst the continuous evaluation of quality is an essential feature of the scientific knowledge production process, the same does not apply to the evaluation of ‘excellence’ and ‘relevance’ for the sake of increasing efficiency and accountability (ibid., 19-20).

Despite this, it is generally accepted that the Higher Education sector is now subject to neoliberal managerial approaches whereby if you can’t count it, it doesn’t count (Feldman & Sandoval, 2018). This increased focus on evaluation to both monitor progress and to incentivise behavioural change has led to a significant increase in the volume of assessments to which researchers, groups and institutions are subject. Whilst over-evaluation is particularly problematic when the range of contributions being assessed is narrow (e.g., the volume, location and citedness of publications (Saenen *et al.*, 2019)) it remains problematic even when applied to a broader diversity of contributions. A new focus on open research practices, integrity and collegiality in our assessments doesn’t displace or even dilute a focus on publications and grant income, but simply expands the number of dimensions on which researchers are assessed. Many of these new dimensions are also not yet mature enough to be evaluated in a robust way at all levels of granularity, which can lead to well-intended but poorly designed evaluations based on limited data.

Poor evaluation design and an overfocus on evaluation for evaluation’s sake has been highlighted as one of the key drivers of many mental health issues in the sector, driving many researchers to seek posts in industry (Gewin, 2022) or worse, to take catastrophic action (Parr, 2014). In response to these considerations, the SCOPE framework urges evaluators to ask at the outset whether they need to evaluate at all, or whether an alternative approach might be taken (such as *enabling* open research practices rather than *evaluating* them). Where an evaluation is deemed necessary, the extent of the evaluation effort should be commensurate with the potential impact of the evaluation. For example, surveys should ask the minimum viable number of questions and the frequency of assessments should be considered carefully.

#### 4.2 Evaluate with the evaluated

The second principle of SCOPE is to evaluate with the evaluated. The principles of co-design are now central to many domains including product and service design, policy design and of course research design itself (Blomkamp, 2018; Moser, 2016; Pinero De Plaza *et al.*, 2023; Steen *et al.*, 2011). Co-design has also been a particularly important tool in efforts to meet equity, diversity and inclusion ambitions (KPMG, 2022). The benefits of co-design are seen to be a more creative process, better outcomes, and greater buy-in by stakeholder communities. It would seem entirely appropriate then, especially given sector concerns about the volume, quality, and format of research assessments, that a principle of co-design and co-evaluation should be adopted.

In this vein, the Leiden Centre for Science and Technology Studies (CWTS) recently proposed ‘evaluative inquiry’ as a more enabling approach to providing material for assessment (Fochler & De Rijcke, 2017). The approach seeks to present research work in ways that represent the complexity of actual practice by engaging with those practitioners instead of taking reductionist approaches to assessment for the sake of standardization (ibid., 34).

The use of SCOPE workshops to develop research evaluations are a good way of ensuring this principle is adhered to.

#### 4.3 Draw on evaluation expertise

The third principle of SCOPE is to draw on evaluation expertise. The ready availability of bibliometric data and tools has led to concern from scientometricians around the rise of ‘armchair bibliometrics’ or ‘citizen scientometricians’ (Leydesdorff *et al.*, 2016). It is common to find academics in every discipline running bibliometric analyses to better understand research activity in their field. This has led others to plead ‘epistemic trespass’. Ballantyne defines epistemic trespassers as “thinkers who have competence or expertise to make good judgements in one field but move to another field where they lack competence – and pass judgement nevertheless” (Ballantyne, 2019, p. 367). When it comes to research assessment, the fact that all researchers are regularly involved in assessing research proposals and applicants for research positions may give them greater confidence that they can expand this knowledge to designing research assessments. More recent emphasis on the need for responsible approaches to research assessment have brought into greater relief how easy it is to get research assessment design wrong. Clearly, the same rigour that is expected of academic research should be expected also of all evaluations of academic research.

### 5. The five-stage SCOPE framework

#### 5.1 Start with what you value

The first stage of SCOPE, *start with what you value*, is a critically important first step in any evaluation. It is about exploring what is valued about the particular entity being evaluated: putting the ‘value’ in e-‘valu’-ation. This approach resonates with the Humane MetricsHSS initiative which supports values-enacted frameworks for evaluating all aspects of



scholarly activity, as well as with Leiden Manifesto's second principle urging that the performance of institutions, groups or researchers should be measured against their missions (Hicks *et al.*, 2015).

An important question when considering what might be valued about an entity under evaluation, is to ask to *whom* the entity offers some value. For example, in a national university research assessment programme there are many stakeholders that all may value different things about the universities being evaluated: the treasury funding the outcomes, the government agency running the evaluation, the institutions themselves, the researchers who work in them, and so on. In line with the SCOPE principle of 'evaluating with the evaluated', the SCOPE approach would be to explore the question across a range of stakeholder perspectives and to seek to find consensus where possible.

If assessments are not developed in line with what stakeholders value about the entity being evaluated, too often they are made in line with what third parties value, or with historical values, or simply in accordance with the data we have available.

The problematic effects of relying on third party values such as the university rankings (see, e.g., Gadd, 2020, 2021; Van Raan, 2005) or national evaluation systems (see, e.g., Aagaard, 2015; Butler, 2005; Watermeyer *et al.*, 2023) are well-documented. The key concern is that given Campbell's Law (what we measure is what we get) (Campbell, 1979), by 'outsourcing' our values to others, evaluators run the risk of producing scholarship and research practices that are not in line with their own mission or ambitions.

The practice of starting with the data that is easily available, and evaluating what can be measured rather than what is valued, is often criticised (Lane *et al.*, 2014). A common focus of such concerns is the over-use of bibliometrics in researcher recruitment and career assessment (Saenen *et al.*, 2021). Proponents of value-led assessment approaches argue that evaluations should not be reduced to the concept of measurable achievements only, as there are multiple contributions that research and researchers make both to scholarship and society (Agate *et al.*, 2020; Holtrop *et al.*, 2020a).

At this stage of SCOPE it is important to maintain the first principle of 'evaluating only where necessary'. Evaluators can fall into the trap of not taking the time to consider what is of the most value and therefore evaluate everything possible.

## 5.2 Consider the context

Discussions around what constitutes a 'good' or 'bad' indicator are abundant in the responsible research evaluation literature (e.g., Rijcke *et al.*, 2016). However, whether an indicator (or indeed any assessment approach) can be determined as suitable or unsuitable depends on the context for that evaluation: what is being measured (entity and discipline) and for what purpose. For this reason, the second stage of SCOPE invites evaluators to *consider the context* of the evaluation.

Entities under evaluation can range from nations through to individuals, and on each level different types of consideration need to be addressed. This is especially the case when those entities are seen through different disciplinary lenses (Holtrop *et al.*, 2020b; Konkiel, 2018; Puuska, 2014; Ylijoki *et al.*, 2011). It is often noted, for example, that whether an evaluation is conducted at a micro- or macro-level significantly affects whether and how quantitative indicators should be used (Waltman, 2018).

In terms of purposes, there are six commonly accepted purposes of research assessment: analysis, advocacy, allocation, accountability, acclaim and adaptation (Parks *et al.*, 2019). The meanings of these terms are often contested by stakeholders, so the INORMS REG have provided short interpretive descriptions of each to aid understanding (see Figure 2). Sometimes evaluations can seek to achieve several different purposes, but it is important to specify these in advance and to consider the purpose in conjunction with the entity under evaluation, in order to ensure the evaluation design is appropriate. What works in one context does not necessarily work in another.

To aid this process, the INORMS REG have developed a matrix plotting the six key evaluation purposes against four different entity sizes to highlight how the impact of an assessment varies (Figure 2). The matrix illustrates that assessments in some settings have more impact on the entity being evaluated and are therefore more 'high risk'. For example, monitoring a country's research performance has less impact on the country being evaluated than evaluating an individual researcher for a promotion, and is therefore arguably a lower risk form of assessment. The exact 'RAG-rating' (red-amber-green) of each of these combinations might be debatable, but the matrix provides a useful heuristic to aid evaluators in understanding the dimensionality of research assessment and to ensuring that assessment approaches are context-sensitive.

		Country	HEI	Group	Individual
Analysis	To understand	Low impact	Low impact	Medium impact	Medium impact
Advocacy	To show off	Low impact	Low impact	Medium impact	Medium impact
Accountability	To monitor	Low impact	Medium impact	Medium impact	High impact
Acclaim	To benchmark	Medium impact	High impact	High impact	High impact
Adaptation	To incentivise	Medium impact	High impact	High impact	High impact
Allocation	To reward	High impact	High impact	High impact	High impact

Low impact  
 Medium impact  
 High impact

**Figure 2. SCOPE ‘Context’ Matrix defining where assessments may have a low/medium/high impact on the assessed entity (from The SCOPE Framework: a five-stage process for evaluating responsibly: <https://doi.org/10.26188/21919527.v1>).**

### 5.3 Options for evaluating

The third stage of SCOPE is to explore all the *options* available for evaluating. This stage is a reminder to consider both quantitative and qualitative approaches and consider them in terms of the values and context of the evaluation (see, e.g., Butler, 2007; De Jong *et al.*, 2011; Gingras, 2014; Holtrop *et al.*, 2020c). The rule of thumb proposed by SCOPE is that quantitative indicators are best reserved for assessing quantitative things: student numbers, money, and citations. In the same way, qualitative approaches are best used for qualitative things: impact and quality. Caution should be taken about using quantitative indicators as a proxy for qualitative things. For example, citation counts are not a suitable proxy for research quality (see, e.g., Aksnes *et al.*, 2019).

It is easy to focus on the dangers of quantitative approaches when considering responsible research assessment, and many declarations and principles do so. However, whilst peer-review is considered the gold standard for research evaluation, it is not without its own challenges (Bornmann & Daniel, 2006; Hug, 2022; Lee *et al.*, 2013; Waltman *et al.*, 2023). Recent concerns about increased journal retractions, and the prevalence of so-called ‘predatory’ journals have raised questions about the quality and reproducibility of peer review. Proponents of open research are calling for greater transparency and openness of peer review, and there are equity, diversity and inclusion concerns (Else & Perkel, 2022).

The truth is that there are limitations to all forms of research assessment, both qualitative and quantitative. For this reason, SCOPE requires evaluators to consider all their assessment options equally. It advocates that in most cases a mixed methods approach is more likely to generate a proportionate and appropriate assessment that will (as with the Hippocratic Oath) first do no harm (Sugimoto & Larivière, 2018). This will always involve human judgement in some form, and always involve an approximation of the reliability of the assessment, through error bars, list of caveats or limitations, and so on.

Given the many and varied values and contexts which may be evaluated, it is not possible to provide a comprehensive list of options for doing so via the SCOPE framework. However, the guidance promotes the use of alternative evaluation approaches such as those provided by the [DORA resource library](#) and the [Metrics Toolkit](#) to offer some inspiration.

At this stage it may be helpful to generate several different options given that the Probe stage will ‘stress-test’ these options and may render some unsuitable. The alternative is to consider both the Options and Probe stages together to ensure that no option is developed to such an extent that it becomes difficult to abandon it after being ‘probed’.

### 5.4 Probe deeply

Once options for evaluating have been selected in line with stakeholder values and context and options, the fourth step is to *probe deeply*. To do this, SCOPE proposes that the evaluator should ask the following four questions of their evaluation:

#### 5.4.1 *Who might the chosen approach discriminate against?*

There is a considerable literature around the biases inherent in all forms of research evaluation as already stated. Demographics most likely fall victim to poor forms of assessment include early-career researchers (Algra *et al.*, 2020), women (Jappelli *et al.*, 2017, 2017; Larivière *et al.*, 2013; Sugimoto & Larivière, 2023), intersectional groups (Bailey, 2018), and those working in non-journal-based disciplines. No evaluation is perfect and, as discussed, there are weaknesses in both qualitative and quantitative forms of assessment. For this reason, it is important to give significant thought to the question as to whether all entities being assessed have equal opportunity to succeed under the evaluation approach selected. If not, what mechanisms might be put in place to mitigate these inherent biases. Much work has been done in this space included the introduction of lotteries to more equitably decide between equally scoring proposals (Roumbanis, 2019), and the use of Unconscious Bias Observers on promotion panels (Bonello *et al.*, 2017). ‘Evaluating with the evaluated’ (SCOPE’s second principle) and ensuring any consulted stakeholder group is representative, will also go some considerable way to addressing this question.

#### 5.4.2 *How might this approach be gamed?*

A mantra often used by the INORMS REG is that ‘where there is a prize there is a game’. This refers to the fact that where there is a lot at stake in a particular evaluation (reputationally and financially), the entities being evaluated will naturally be incentivised to alter their behaviours in ways that enable them to perform well (Biagioli & Lippman, 2020). In some cases, this is indeed the purpose of an evaluation: evaluators will seek to assess a particular dimension (e.g., open research) to incentivise it. However, there is a spectrum of responses from evaluated parties to evaluation efforts, from legitimately optimising their activities (e.g., making more outputs open access), to ‘gaming’ their submissions (e.g., only reporting outputs once they’ve been made open access), to outright cheating (fabricating open access data). A strong evaluation should seek to anticipate potential opportunities for gaming with a view to designing them out of the system.

#### 5.4.3 *What might the unintended consequences be?*

In his book, *The Tyranny of Metrics*, Muller devotes a whole chapter to “the unintended but predictable negative consequences” of poor assessment practices and indicators (Muller, 2018). He describes some common unintended consequences in terms of goal displacement, short-termism, diminishing utility, rewarding luck, and discouraging risk-taking, innovation and cooperation. Trying to predict the potential harmful consequences of an evaluation approach into which the evaluator has invested much care and effort, is a difficult ask. Whilst it is not always possible to predict unintended consequences, it is important to attempt to do so at both an institutional as well as an individual level (see, e.g., Dahler-Larsen, 2014; Lorenz, 2014; Rijke *et al.*, 2016; Stephan *et al.*, 2017; Wellcome Trust, 2020). The use of workshops to design evaluations ‘with the evaluated’, where the evaluated act as ‘critical friends’ is one useful way of identifying some of the unintended consequences before time and expense is invested into running the evaluation. However, it should be accepted that some consequences are not always predictable, and this is a question that should be returned to at the ‘Evaluate’ stage of SCOPE.

#### 5.4.4 *Does the cost of measuring outweigh the benefit?*

Another of Mueller’s unintended consequences of evaluation is the significant costs that may be incurred in both running and interpreting the assessment (Muller, 2018). SCOPE is clear that the cost, including the workload, stress, and finances, should be proportional to the aims and anticipated outcomes of the evaluation (Sawczak, 2018). Given the strongest evaluations usually consist of a mixed methods approach involving some element of human judgement, the cost of an evaluation can quickly escalate.

A current case in point is the reported cost of running the 2021 UK Research Excellence Framework which came in at £471 million (Research England *et al.*, 2023). This is only 3–4% of the block-grant funding linked to its outcomes; however, it is almost double the cost of the 2014 exercise (Else, 2015), which was in turn three times higher than the 2008 exercise (Sayer, 2015). The Joint UK HE Funding bodies have explicitly stated an intention to reduce the cost of the exercise in 2028. Similarly, the Danish government recently announced that they would discontinue the updating of the national bibliometric research indicator due to the cost and burden not being commensurate with the benefit (Uddannelses- og Forskningsministeriet, 2021).

There is no ‘correct answer’ when it comes to the cost:benefit ratio of an evaluation. However, as with all business decisions, those making the investment need to be reassured of an appropriate return. While evaluations can be extremely beneficial to generate intelligence, evidence, improve efficiencies or identify gaps, these benefits are only realized if the evaluation is designed in a way that provides usable outcomes at a reasonable cost.

## 5.5 Evaluate your evaluation

The fifth, and final stage of SCOPE is to *evaluate your evaluation*. After conducting an evaluation, it is important to check if it reached its aims, if the results are useful, and if the evaluation approach brought new insights to what was being evaluated. Did the evaluation cause unintended consequences not foreseen at the Probe stage? If so, they should be considered when interpreting the results and addressed prior to future assessments. Even where an approach proved to be successful, it should be kept in mind that the tools available to undertake an evaluation (e.g., the data sources and indicators available) as well as values, missions, and strategies, are subject to change.

This last step of SCOPE is often overlooked but is of vital importance. Patton & Campbell-Patton consider evaluations meaningless, unless their findings are used (Patton & Campbell-Patton, 2022), and Stufflebeam considers meta-evaluation – the evaluation of evaluations – as a professional obligation of evaluators (Stufflebeam, 2001; see also, Scriven, 2009). In addition to any immediate post-assessment evaluation, established evaluations should also be re-evaluated at regular intervals to ensure they are still in alignment with what is valued about the entity under evaluation, and does not result in any unintended consequences that may require an adjustment to the evaluation. One of the strengths of SCOPE is that it can be used to both design new evaluations, and to assess existing evaluations. Thus the ‘E’ of SCOPE is really an invitation to run through the SCOPE process again to assess the evaluation that has been designed and implemented.

Whilst SCOPE is presented as a linear, step-by-step, process, it is rather more iterative in practice. Each stage of SCOPE might send the evaluator back to a previous stage to reconsider a prior decision. For example, the unintended consequences unearthed at the Probe stage might cause the evaluator to consider whether the Options they have chosen are sensible, or even whether the evaluation is suitable for a particular Context. Ultimately, as long as each stage of SCOPE is considered in the design of an evaluation, it has a strong chance of being an appropriate and proportionate assessment.

## 6. Use cases

The SCOPE framework is in wide usage by a range of organisations globally. Recent examples include the use of SCOPE by Indian funding agencies to develop assessment mechanisms (Suchiradipta *et al.*, 2023), by Colombian research professionals to develop a responsible assessment policy (Pallares *et al.*, 2023), by Finnish research managers to create an institutional policy on assessment of researchers (University of Turku, n.d.) and by UK research leaders to develop an approach to assessing research culture (Davies & Fadhel, 2023). Example case studies are regularly added to the [SCOPE webpage](#). This section provides four case studies to demonstrate how it has been used by different types of organisations to develop evaluations at different levels of granularity and in different settings.

### 6.1 Case Study: Emerald Publishing, UK

#### 6.1.1 Background

Emerald Publishing is a global scholarly publisher committed to equity, diversity, and inclusion. As such they were keen to start monitoring and incentivising greater diversity and representation on the editorial boards of their scholarly journals. Being aware of the sensitivities around this, they used the SCOPE framework to explore how they might do so (Gadd & Himanen, 2021a).

#### 6.1.2 Process

The publisher firstly ran a 90-minute *Start with what you value* workshop with ten editors and editorial board members and supported by the INORMS REG. This explored ‘what might a diverse editorial board look like’ and ‘what aspects of diversity actually benefit a journal’ in the agreed *context* of ‘incentivising’ greater diversity in editorial boards. Given the SCOPE principle of evaluating only where necessary, a discussion was held as to whether the valued dimensions of diversity should be enabled rather than evaluated. *Options* for evaluating this value in these contexts were discussed with Emerald first presenting some examples of what can be assessed using existing systems and approaches.

A second 90-minute workshop was then run to explore *Options* in greater detail with the Emerald Publishing Editorial and Rights/Legal team members. The options were *probed* using the four key *probe* questions at the same time.

#### 6.1.3 Outcomes

A key learning point from the *values* stage was that the editorial board members’ views of diversity included subject diversity, diversity in the way knowledge is/can be disseminated as well as regional diversity. Editorial Board members also felt that a commitment to diversity should be taken as part of a wider review of editorial board processes and policies,

e.g., dormancy in editorial boards and whether boards would benefit from shorter-term appointments. Also, they felt that diversity within editorial boards was necessarily affected by the broader challenges around the prevailing model of creating scientific knowledge (e.g., the use of unpaid editorial positions) and global inequality more broadly.

Considering the *context* of ‘incentivisation’ led to interesting discussions as to whether this value needed enabling or evaluating. Interestingly, the workshop members felt it needed both, and that some sort of ‘badges’ or external signifier that the journal was at least committed to improving diversity, would be welcomed. Another important question asked who was being incentivised in this context: Emerald Publishing, editors and/or editorial board members? Ultimately it was agreed that editors should be the focus of any enabling and evaluating activity, but the relationship between publisher and editors is a carefully balanced one.

While a mixed evaluate-and-enable approach was agreed upon, many of the actions and ideas that came from the workshops were focussed on enabling diversity. *Options* floated included both paying editorial board members to allow less-well-funded scholars to participate and introducing a Diversity Editorial Pledge whereby editors could be rewarded for a commitment to diversity. *Probing* these *options* generated questions around the sensitivities and challenges associated with monitoring diversity-related data.

The SCOPE process resulted in a range of short, medium, and long-term actions for Emerald Publishing to pursue. The immediately actionable items included building expectations around diversity into contracts/job descriptions for Editors; making diversity a rolling agenda point for Editorial meetings and reporting on progress around diversity in Editorial Advisory Board (EAB) meetings; and encouraging editors and EABs to identify their own success EDI indicators in line with their values.

## 6.2 Case Study: The UK HE Joint Funding Bodies ‘Future Research Assessment Programme’

### 6.2.1 Background

Research England, the Scottish Funding Council, the Northern Ireland Department for the Economy, and the Higher Education Funding Council for Wales commissioned the Future Research Assessment Programme (FRAP) with a view to designing a new national research assessment exercise in the UK. The outcomes of the current system are used to inform the allocation of quality-related research funding to UK universities and provide accountability for public investment in research. It was felt that without a proper framework such a review could focus narrowly on making minor adjustments to the existing system (‘evolution’) or solely rely on learning from other international research assessment approaches (‘reproduction’) when what the funding bodies really sought was a root-and-branch review (possible ‘revolution’). It was felt that the SCOPE framework offered a structure for thinking about such a review, providing “the perfect brief to be radical” (Gadd & Himanen, 2021b, p. 2).

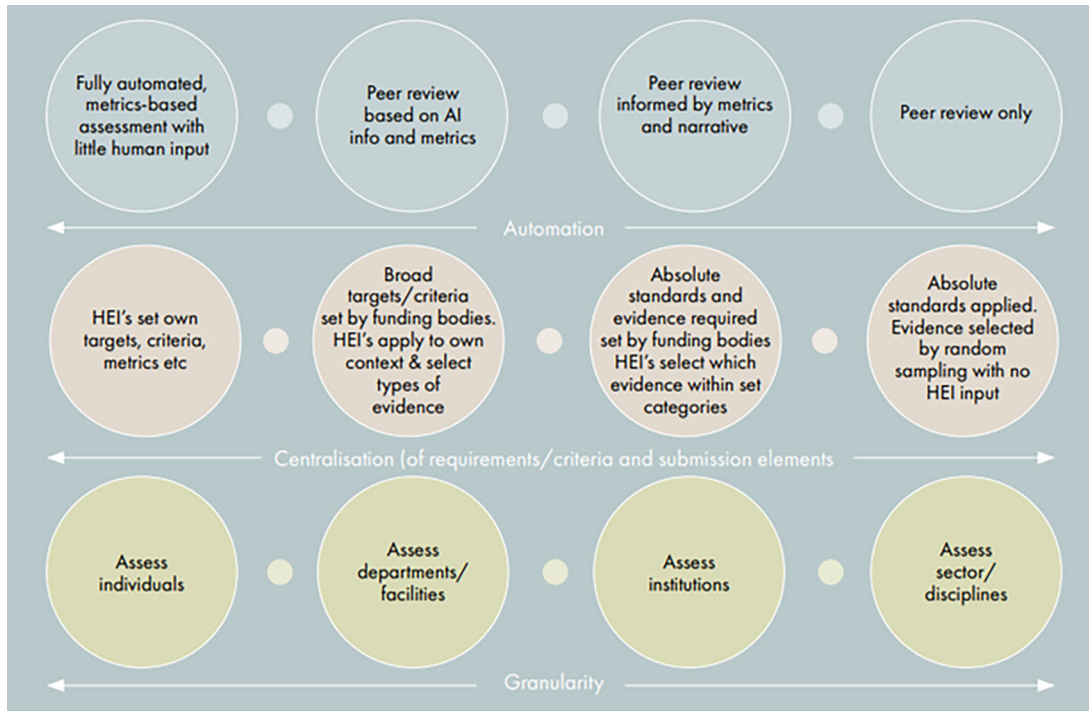
### 6.2.2 Process

The use of SCOPE by the funding bodies was unmediated by the INORMS REG (Gadd & Himanen, 2021b). The FRAP team were keen to *start with what was valued* about the research system. Deploying the ‘Evaluate with the evaluated’ principle, they held a series of round-table events with different stakeholder groups. To create a healthy ecosystem, it was important to the team to not only start with what was valued, but to also agree which of those valued things should be evaluated. Understanding the *contexts* in which the national research evaluation exercise should take place was an important step in the process. The REF exercise that ran in 2021 served a wide range of purposes and the team sought to identify which were non-negotiable. The FRAP team were inspired by the *options* stage of exploring both qualitative and quantitative options to develop a set of spectra to understand the community’s appetite for different variables, for example, around the degree of automation, centralisation, granularity, and frequency (see Figure 3).

Of particular interest to the funding bodies was to *probe* for unintended consequences and perverse incentives. The cost-benefit of the exercise was explored in a dedicated assessment (Neto *et al.*, 2023).

### 6.2.3 Outcomes

The REF 2028 Initial Decisions (Research England *et al.*, 2023) show many of the hallmarks of a SCOPE-led assessment. The principle of ‘evaluating with the evaluated’ and starting with what was *valued* is clearly in evidence and the principle of drawing on evaluation expertise’ was fulfilled by utilising specialists to produce reports on metrics (Curry *et al.*, 2022), artificial intelligence (Thelwall *et al.*, 2022) and a ‘real-time REF review’ to support the process (Manville *et al.*, 2021). The clear articulation of the purposes of the next REF specifies the *context* and the *options* carefully balance both



**Figure 3.** Spectra used to identify stakeholder appetites for different evaluation variables (from INORMS SCOPE Case Study: The UK Higher Education Funding Bodies - Developing a new national research evaluation system: <https://inorms.net/wp-content/uploads/2021/11/inorms-scope-case-study-uk-he-funding-bodies-final.pdf>).

qualitative and quantitative measures in an effort to reduce burden. The Initial Decisions document makes frequent reference to the need to mitigate unintended consequences particularly to under-represented groups. Further consultation is afoot, again on the principle of ‘Evaluating with the Evaluated’ to ensure no foreseeable but unintended consequences are at play.

### 6.3 Case Study: Newcastle University, UK<sup>1</sup>

#### 6.3.1 Background

To support Newcastle University’s work on enhancing their research culture, they used the SCOPE framework to develop a set of research culture Key Performance Indicators (KPIs) for the University’s Research Strategy and a basket of measures for their ‘Research Culture Index’.

#### 6.3.2 Process

Using the SCOPE framework and supported by the INORMS REG, the University held an initial community workshop with around 80 colleagues (both academic and professional services colleagues) and postgraduate students to identify what people *valued* in a positive research culture. They sought to understand both what a positive research culture ‘looked and felt like’ to fully understand how it might enhance the research community’s experiences in relation to doing great research. The resulting values were then tested and refined at a subsequent smaller workshop of some of the attendees.

Once agreed, a smaller facilitated focus group was held to generate a series of *options* of measures for evaluation within the *context* of ‘monitoring’ as KPIs. The options were then *probed* for any unintended consequences. The resulting set of measures were shared with colleagues with relevant expertise within the University to check for feasibility and viability in

<sup>1</sup>The case description is based on INORMS REG members’ meetings with Sarah Whalley, Research Strategy Manager at Newcastle University.

the *context* of different Faculty-based disciplines and given the practicalities of the University's research reporting systems.

### 6.3.3 Outcomes

The first high-level workshop identified four key attributes of a positive research culture in terms of:

- Collaboration and collegiality
- The freedom to explore and grow
- Fairness and inclusion
- Openness and integrity

The second workshop explored the dimensions (or 'sub-values') of these attributes that constituted the 'look and feel' of a positive research culture. These included strong support for the careers of others, a sense of belonging, and increasing empowerment and satisfaction. These sub-values more easily lent themselves to the development of specific *options* for monitoring improvement.

Seventeen options were ultimately selected after a number were excluded during the *probe* stage. One example of an excluded measure was 'Bullying and harassment reporting.' This was excluded as it would only reflect reporting rather than occurrence, and it was unclear whether the measure should increase or decrease. For example, an increase in reporting could be interpreted as a positive if people feel safer and more supported to report, but also a negative if it captures increasing incidence. The final outcome of the process was a 'Research Culture Index' with seventeen dimensions and some of these informed the revision of the University's Research Strategy KPIs.

## 6.4 Case Study: University of Alberta, Canada<sup>2</sup>

### 6.4.1 Background

The University of Alberta, in Edmonton, Canada, has been exploring ways to implement responsible research evaluation approaches throughout the University. Particularly their Research Impact Librarians were interested in learning how to conduct research assessment more effectively and more equitably. Upon learning of the SCOPE framework, they identified the need to provide training on this approach to various members of their community.

### 6.4.2 Process

It was determined to hold two consecutive workshops assisted by the INORMS REG that focused on the exploration and then implementation of the SCOPE framework, with the goal to increase capacity within the University of Alberta to develop, evaluate and refine responsible research assessments.

The first workshop was more didactic in nature as many attendees were not yet familiar with the SCOPE framework. As such, this 90-minute session focused on the principles and stages within the SCOPE framework, providing the knowledge and understanding behind each stage of the process. This workshop was attended by a broad range of individuals including senior university leadership, representatives from various faculties and departments, and library information specialists.

The second 2-hour workshop focused on the direct implementation and application of the SCOPE framework with a smaller subset of individuals who routinely develop or participate in research assessment. Since the University was working to develop better assessment of knowledge mobilisation (KM), this was used as the working example. For each stage of the SCOPE framework various small group exercises or group discussions were held to explore the aspects of KM that were *valued* the most (impact in policy, uptake by community, scientific knowledge, etc.); in which *context* they sought to assess KM (advocacy of the organization, accountability of departments, etc.); what were the *options* they had, or could develop, to assess KM (collaborations, policy changes, publications, commercialisation, etc.); how they could

---

<sup>2</sup>The case description is based on INORMS REG members' meetings with Thane Chambers, Librarian at University of Alberta.

*probe* these options to determine if there was bias or discrimination that needs to be considered or addressed, and finally how they could thoughtfully *evaluate* their assessment to determine if it met the principles of the SCOPE framework.

#### 6.4.3 Outcome

The University of Alberta found these workshops increased both knowledge and capacity in the responsible research assessment. While KM was used as a case example during the workshop, it was acknowledged that development of robust framework would need to be co-developed with those being evaluated and should also involve reaching out to others who were heavily invested in the assessment of KM, such as Research Impact Canada, to draw on available expertise. The University of Alberta now has plans to use SCOPE in ongoing development of multiple assessments and in the re-evaluation of previously established processes.

## 7. Discussion and conclusions

The SCOPE framework seeks to support evaluators in any and every research setting to implement the many complementary principles of RRA in the design and delivery of their evaluation approaches. In this way it bridges the gap between principles and practice. However, SCOPE does not simply translate existing principles into practice, but provides additional considerations not always addressed by principles of RRA, such as evaluating with the evaluated, evaluating only where necessary, starting with what is valued and probing for unintended consequences. Thus SCOPE, with its focus on implementation, plays a specific role in the delivery of RRA.

The use cases presented show the value of SCOPE across a range of settings. This highlights one of the strengths of SCOPE, namely, that it is widely applicable and enables both quantitative and qualitative assessments for any purpose, at any level, and any discipline and can be used by any evaluator with any background (assuming of course, that they draw on appropriate expertise as required by principle three). To develop a framework with such wide applicability has necessitated it to take a very high-level, somewhat simplified, approach to research assessment. Indeed, the whole framework can be presented in a one-page overview (see [Figure 1](#)). This is both a strength and a weakness.

It is a strength in the sense that the main principles and stages can be communicated and understood in a few minutes. It can also be used with or without assistance from the INORMS REG as evident from the use cases. This is probably why it has captured the imagination of the global research community and been so widely adopted. However, as can be seen by the full SCOPE guide there is a lot more underneath the simple heuristics to be explored and understood ([International Network Of Research Management Societies-Research Evaluation Group, 2021](#)), which is where the weakness lies: an evaluator that has sought to apply SCOPE without drawing on appropriate evaluation expertise may misinterpret some of the steps and claim they have a ‘SCOPE-compliant’ evaluation where this might not be the case.

When given five seemingly simple steps, it is possible to just follow them mechanically, “tick the box”, without giving enough consideration to what the steps are there for. Conversely it may be difficult to discern when you have completed a step of the framework: How do you know when you have accurately identified your values? When have you probed enough into unintended consequences? As a framework SCOPE leaves this up to the discretion of the evaluating group as time frames and amount of engagement may vary based on the organization and topic being evaluated.

Other feedback sometimes received is that the framework is common sense and aligned with existing practice. Whilst the team would agree with the former, continued evidence of poorly designed research assessments gives the lie to the latter. SCOPE is simple but is not universally applied. Were each of the stages of SCOPE properly applied under its three principles, the existence of problematic research evaluations would be greatly diminished. What might look like common sense at first is, in reality, a series of deep and fundamental questions enabling both the evaluators and the evaluated to systematically reflect on their practices and their assumptions. This reflection leads to an explicit definition of the values that are the foundation for an evaluation and focuses attention on potential biases and weaknesses in the evaluation design, which may not ordinarily be given their due attention. Thus, where evaluation may previously have been an implicit part of publishing, hiring, budgeting, etc., SCOPE puts a focus on evaluation in its own right.

One of the strengths of SCOPE is that in addition to enabling the design of responsible research assessments, it can act as a training framework for research evaluators. Many more professionals (research managers, planners, funders, librarians, and publishers) are being called upon to evaluate research with no formal training, nor the capacity or opportunity to undertake appropriate training. Many researchers who have had to participate in some form of evaluation as part of their roles (journal peer review, recruitment, etc.) might feel that this knowledge is transferable to other forms of evaluation without recognising some of the differences between various forms of assessment. SCOPE is simple enough and accessible enough to provide a framework for a deeper understanding of responsible research assessment practice and



could support the greater professionalisation of research evaluation. However, one obvious limitation of SCOPE is its lack of more concrete instructions on which tools or indicators should be used to ensure a responsible assessment. Such recommendations were beyond the scope of the framework, mainly due to the plethora of different approaches available, and the complicated relationship they have with different contexts and purposes of evaluation. However, this can leave evaluators at a loss to know whether they have fully explored all of the options available to them. The current paper focuses on placing SCOPE in the landscape of RRA principles and points to justifications for the SCOPE principles and stages in scholarly literature. Having said that, an application that supports the selection of appropriate options for evaluating is something that could be considered in the future.

As demonstrated in this paper each of the three principles and the five stages in the SCOPE framework are rooted firmly in the existing research literature, whilst consolidating and expanding on this evidence with lessons learned from experience. Whilst RRA began as a series of objections against data-driven research evaluations, the SCOPE framework provides evaluators with a more positive, comprehensive, and practical approach to all forms of research assessment. It is offered up to the community as a useful tool in the toolbox of all research evaluators.

### Author contributions

- Conceptualization: all
- Investigation: all
- Methodology: all
- Project administration: Himanen, Gauffriau, Gadd
- Supervision: Gadd
- Writing – original draft: Himanen, Conte, Gauffriau, Gadd
- Writing – review & editing: all

### Data availability

All data underlying the results are available as part of the article and no additional source data are required.

### Acknowledgements

The authors are grateful for the contributions to the four case studies from Sally Wilson of Emerald Publishing, Catriona Firth of Research England, Candy Rowe & Sarah Whalley of Newcastle University, and Thane Chambers of the University of Alberta.

### References

- Aagaard K: **How incentives trickle down: Local use of a national bibliometric indicator system.** *Sci. Public Policy.* 2015; **42**(5): 725–737.  
[Publisher Full Text](#)
- Agate N, Kennison R, Konkiel S, *et al.*: **The transformative power of values-enacted scholarship.** *Humanities and Social Sciences Communications.* 2020; **7**(1): 165.  
[Publisher Full Text](#)
- Aksnes DW, Langfeldt L, Wouters P: **Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories.** *SAGE Open.* 2019; **9**(1): 215824401982957.  
[Publisher Full Text](#)
- Algra A, Koopman I, Snoek R: **How young researchers can re-shape the evaluation of their work.** *Nature Index News.* 2020.  
[Reference Source](#)
- Aubert Bonn N, De Vries RG, Pinxten W: **The failure of success: Four lessons learned in five years of research on research integrity and research assessments.** *BMC. Res. Notes.* 2022; **15**(1): 309.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Aubert Bonn N, Pinxten W: **Advancing science or advancing careers? Researchers' opinions on success indicators.** *PLoS One.* 2021a; **16**(2): e0243664.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Aubert Bonn N, Pinxten W: **Rethinking success, integrity, and culture in research (part 1)—A multi-actor qualitative study on success in science.** *Research Integrity and Peer Review.* 2021b; **6**(1): 1.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bailey M: **On misogynoir: Citation, erasure, and plagiarism.** *Fem. Media Stud.* 2018; **18**(4): 762–768.  
[Publisher Full Text](#)
- Ballantyne N: **Epistemic Trespassing.** *Mind.* 2019; **128**(510): 367–395.  
[Publisher Full Text](#)
- Benedictus R, Miedema F, Ferguson MWJ: **Fewer numbers, better science.** *Nature.* 2016; **538**(7626): 453–455.  
[Publisher Full Text](#)
- Biagioli M, Lippman A: *Gaming the metrics: Misconduct and manipulation in academic research.* The MIT Press; 2020.  
[Publisher Full Text](#)
- Birukou A, Wakeling JR, Bartolini C, *et al.*: **Alternatives to Peer Review: Novel Approaches for Research Evaluation.** *Front. Comput. Neurosci.*

2011; 5.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)Blomkamp E: **The Promise of Co-Design for Public Policy.** *Aust. J. Public Adm.* 2018; **77**(4): 729–743.[Publisher Full Text](#)Bonello E, Coombs H, Dessent C: **Unconscious bias observer scheme: DEVELOPMENT OF A NOVEL UNCONSCIOUS BIAS OBSERVER SCHEME AT CHEMISTRY@YORK.** Sheffield: LGBSTEMinar; 2017.[Reference Source](#)Bornmann L, Daniel H-D: **Potential sources of bias in research fellowship assessments: Effects of university prestige and field of study.** *Research Evaluation.* 2006; **15**(3): 209–219.[Publisher Full Text](#)Bornmann L, Marewski JN: **Heuristics as conceptual lens for understanding and studying the usage of bibliometrics in research evaluation.** *Scientometrics.* 2019; **120**(2): 419–459.[Publisher Full Text](#)Butler L: **What Happens When Funding is Linked to Publication Counts?** Moed HF, Glänzel W, Schmoch U, editors. *Handbook of Quantitative Science and Technology Research.* Kluwer Academic Publishers; 2005; (pp. 389–405).[Publisher Full Text](#)Butler L: **Assessing university research: A plea for a balanced approach.** *Sci. Public Policy.* 2007; **34**(8): 565–574.[Publisher Full Text](#)Campbell DT: **Assessing the impact of planned social change.** *Eval. Program Plann.* 1979; **2**(1): 67–90.[Publisher Full Text](#)Curry S, Gadd E, Wilsdon J: **Harnessing the Metric Tide: Indicators, infrastructures & priorities for UK responsible research assessment.** Research on Research Institute; 2022; 11014215 Bytes.[Publisher Full Text](#)Curry S, de Rijcke S, Hatch A, et al.: **The changing role of funders in responsible research assessment: Progress, obstacles and the way ahead (RoRI Working Paper No.3) (p. 2449096 Bytes).** *Research on Research Institute.* 2020.[Publisher Full Text](#)Dahler-Larsen P: **Constitutive Effects of Performance Indicators: Getting beyond unintended consequences.** *Public Manag. Rev.* 2014; **16**(7): 969–986.[Publisher Full Text](#)Davies C, Fadhel S: **Using SCOPE to select metrics to track improvements in Research Culture: Interim reflections.** University of Leeds; 2023.[Reference Source](#)De Jong SPL, Van Arensbergen P, Daemen F, et al.: **Evaluation of research in context: An approach and two cases.** *Research Evaluation.* 2011; **20**(1): 61–72.[Publisher Full Text](#)Else H: **REF 2014 cost almost £250 million.** *Times Higher Education.* 2015.[Reference Source](#)Else H, Perkel JM: **The giant plan to track diversity in research journals.** *Nature.* 2022; **602**(7898): 566–570.[PubMed Abstract](#) | [Publisher Full Text](#)European University Association, Science Europe, European CommissionStroobants K: **Agreement on reforming research assessment.** 2022; (p. 23).[Reference Source](#)Feldman Z, Sandoval M: **Metric Power and the Academic Self: Neoliberalism, Knowledge and Resistance in the British University.** *TripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society.* 2018; **16**(1): 214–233.[Publisher Full Text](#)Fochler M, De Rijcke S: **Implicated in the Indicator Game? An Experimental Debate.** *Engag. Sci. Technol. Soc.* 2017; **3**: 21–40.[Publisher Full Text](#)Gadd E: **University rankings need a rethink.** *Nature.* 2020; **587**(7835): 523–523.[PubMed Abstract](#) | [Publisher Full Text](#)Gadd E: **Mis-measuring our universities: Why global university rankings don't add up [Preprint].** *SocArXiv.* 2021.[Publisher Full Text](#)Gadd E, Himanen L: **INORMS SCOPE Case Study: Emerald Publishing: Evaluating the diversity of editorial boards.** 2021a; (p. 5).[Reference Source](#)Gadd E, Himanen L: **INORMS SCOPE Case Study: The UK Higher Education Funding Bodies: Developing a new national research evaluation system.** 2021b; (p. 5).[Reference Source](#)Gauffriau M: **Navigating Responsible Research Assessment Guidelines.** *Leiden Madrics.* 2023.[Reference Source](#)Geuna A, Martin BR: **University Research Evaluation and Funding: An International Comparison.** *Minerva.* 2003; **41**(4): 277–304.[Publisher Full Text](#)Gewin V: **Has the 'great resignation' hit academia?.** *Nature.* 2022; **606**(7912): 211–213.[PubMed Abstract](#) | [Publisher Full Text](#)Gingras Y: **Criteria for evaluating indicators.** *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact.* The MIT Press; 2014; (pp. 109–125).Gläser J, Laudel G: **Evaluation Without Evaluators.** Whitley R, Gläser J, editors. *The Changing Governance of the Sciences.* Netherlands: Springer; 2007; Vol. **26**: pp. 127–151.[Publisher Full Text](#)Hallonsten O: **Stop evaluating science: A historical-sociological argument.** *Soc. Sci. Inf.* 2021; **60**(1): 7–26.[Publisher Full Text](#)Hicks D, Wouters P, Waltman L, et al.: **Bibliometrics: The Leiden Manifesto for research metrics.** *Nature.* 2015; **520**(7548): 429–431.[PubMed Abstract](#) | [Publisher Full Text](#)Holtrup T, Hessels L, Prins A: **Evaluative Inquiry I: Academic value is more than performance.** *Leiden Madrics.* 2020a.[Reference Source](#)Holtrup T, Hessels L, Prins A: **Evaluative Inquiry II: Evaluating research in context.** *Leiden Madrics.* 2020b.[Reference Source](#)Holtrup T, Hessels L, Prins A: **Evaluative Inquiry III: Mixing methods for evaluating research.** *Leiden Madrics.* 2020c.[Reference Source](#)Hug SE: **Towards theorizing peer review.** *Quant. Sci. Stud.* 2022; **3**(3): 815–831.[Publisher Full Text](#)International Network Of Research Management Societies-Research Evaluation Group: **The SCOPE Framework: A five-stage process for evaluating research responsibly.** International Network of Research Management Societies; 2021; 668558 Bytes.[Publisher Full Text](#)Jappelli T, Nappi CA, Torrini R: **Gender effects in research evaluation.** *Res. Policy.* 2017; **46**(5): 911–924.[Publisher Full Text](#)Konkiel S: **Approaches to creating 'humane' research evaluation metrics for the humanities.** *Insights the UKSG Journal.* 2018; **31**: 44.[Publisher Full Text](#)KPMG: **Health workforce equity, diversity and inclusion: Taking deliberate actions to develop inclusive and equitable workplace cultures.** 2022.[Reference Source](#)Lamont M, Huutoniemi K: **Comparing Customary Rules of Fairness: Evaluative Practices in Various Types of Peer Review Panels.** Camic C, Gross N, Lamont M, editors. *Social knowledge in the making.* University of Chicago Press; 2011.Lane J, Largent M, Rosen R: **Science Metrics and Science Policy.** Cronin B, Sugimoto CR, editors. *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact.* The MIT Press; 2014; (pp. 397–411).Larivière V, Ni C, Gingras Y, et al.: **Bibliometrics: Global gender disparities in science.** *Nature.* 2013; **504**(7479): 211–213.[PubMed Abstract](#) | [Publisher Full Text](#)Lebel J, McLean R: **A better measure of research from the global south.** *Nature.* 2018; **559**(7712): 23–26.[Publisher Full Text](#)Lee CJ, Sugimoto CR, Zhang G, et al.: **Bias in peer review.** *J. Am. Soc. Inf. Sci. Technol.* 2013; **64**(1): 2–17.[Publisher Full Text](#)Leydesdorff L, Wouters P, Bornmann L: **Professional and citizen bibliometrics: Complementarities and ambivalences in the development and use of indicators—a state-of-the-art report.** *Scientometrics.* 2016; **109**(3): 2129–2150.[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)Lorenz C: **Fixing the facts: The rise of new public management, the metrification of "quality" and the fall of the academic professions.** *Moving the Social.* 2014; **52**: 5–26.Manville C, d'Angelo C, Culora A, et al.: **Understanding perceptions of the Research Excellence Framework among UK researchers: The Real-Time REF Review.** RAND Corporation; 2021; 140.[Reference Source](#)Merton RK: **The normative structure of science.** Merton RK, Storer NW, editors. *The Sociology of Science: Theoretical and Empirical Investigations.* The University of Chicago Press; 1973.Moher D, Bouter L, Kleinert S, et al.: **The Hong Kong Principles for assessing researchers: Fostering research integrity.** *PLoS Biol.* 2020; **18**(7): e3000737.[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

- Moher D, Naudet F, Cristea IA, et al.: **Assessing scientists for hiring, promotion, and tenure.** *PLoS Biol.* 2018; **16**(3): e2004089.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Morrish L: **Why the audit culture made me quit.** *Times Higher Education.* 2017.  
[Reference Source](#)
- Moser SC: **Can science on transformation transform science? Lessons from co-design.** *Curr. Opin. Environ. Sustain.* 2016; **20**: 106–115.  
[Publisher Full Text](#)
- Muller J: *The Tyranny of Metrics.* Princeton University Press; 2018.  
[Publisher Full Text](#)
- Neto A, Vingre A, D'Hont J, et al.: **REF 2021 Cost Evaluation: Final report.** 2023; (p. 78).  
[Reference Source](#)
- Pallares C, Chalela S, Tejada MA, et al.: **The Colombian responsible metrics Project: Towards a Colombian institutional, methodological instrument for research assessment.** *DORA Blog.* 2023.  
[Reference Source](#)
- Parks S, Rodriguez-Rincon D, Parkinson S, et al.: *The changing research landscape and reflections on national research assessment in the future (RR-3200-UKRI).* RAND Corporation; 2019; 204.  
[Reference Source](#)
- Parr C: **Imperial College professor Stefan Grimm 'was given grant income target.'** *Times Higher Education.* 2014.  
[Reference Source](#)
- Patton MQ, Campbell-Patton CE: *Utilization-focused evaluation.* 5th edition. SAGE; 2022.
- Pinero De Plaza MA, Yadav L, Kitson A: **Co-designing, measuring, and optimizing innovations and solutions within complex adaptive health systems.** *Front. Health Serv.* 2023; **3**: 1154614.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Puuska H-M: *Scholarly Publishing Patterns in Finland: A comparison of disciplinary groups.* Tampere University; 2014.  
[Reference Source](#)
- Research England, Scottish Funding Council, Higher Education Funding Council for Wales, & Department for the Economy, Northern Ireland: **Research Excellence Framework 2028: Initial decisions and issues for further consultation (REF 2028/23/01; p. 34).** 2023.  
[Reference Source](#)
- Rijcke SD, Wouters PF, Rushforth AD, et al.: **Evaluation practices and effects of indicator use—A literature review.** *Research Evaluation.* 2016; **25**(2): 161–169.  
[Publisher Full Text](#)
- Rossi PH, Lipsey MW, Henry GT: *Evaluation: A systematic approach.* Eighth edition. SAGE; 2019.
- Roumbanis L: **Peer Review or Lottery? A Critical Analysis of Two Different Forms of Decision-making Mechanisms for Allocation of Research Grants.** *Sci. Technol. Hum. Values.* 2019; **44**(6): 994–1019.  
[Publisher Full Text](#)
- Rushforth A, Hammarfelt B: **The rise of responsible metrics as a professional reform movement: A collective action frames account.** *Quant. Sci. Stud.* 2023; **4**(4): 879–897.  
[Publisher Full Text](#)
- Saenen B, Hatch A, Curry S, et al.: *Reimagining Academic Career Assessment: Stories of innovation and change.* DORA; European University Association; SPARC Europe; 2021; 47.  
[Reference Source](#)
- Saenen B, Morais R, Gaillard V, et al.: *Research Assessment in the Transition to Open Science: 2019 EUA Open Science and Access—Survey Results.* European University Association; 2019; 48.  
[Reference Source](#)
- San Francisco Declaration on Research Assessment: 2012.  
[Reference Source](#)
- Sawczak K: **The hidden costs of research assessment exercises: The curious case of Australia.** *LSE Impact Blog.* 2018.  
[Reference Source](#)
- Sayer D: **Why did REF2014 cost three times as much as the RAE? Hint: It's not just because of the added impact element.** *LSE Impact Blog.* 2015.  
[Reference Source](#)
- Scriven M: **Meta-Evaluation Revisited.** *Journal of MultiDisciplinary Evaluatio.* 2009; **6**(11): iii–viii.
- Steen M, Manschot M, De Koning N: **Benefits of Co-design in Service Design Projects.** *Int. J. Des.* 2011; **5**(2): 53–60.
- Stephan P, Veugelers R, Wang J: **Reviewers are blinkered by bibliometrics.** *Nature.* 2017; **544**(7651): 411–412.  
[Publisher Full Text](#)
- Stufflebeam DL: **The Metaevaluation Imperative.** *Am. J. Eval.* 2001; **22**(2): 183–209.  
[Publisher Full Text](#)
- Suchiradipta B, Koley M, Bharadwa J: **Workshop on Research Assessment Practices in Indian Funding Agencies.** *Journal of Science Policy & Governance.* 2023; **22**(1).  
[Publisher Full Text](#)
- Sugimoto CR, Larivière V: *Measuring research: What everyone needs to know.* Oxford University Press; 2018.
- Sugimoto CR, Larivière V: *Equity for women in science: Dismantling systemic barriers to advancement.* Harvard University Press; 2023.
- Thelwall M, Kousha K, Wouters P, et al.: **The metric tide: Literature review.** 2015.  
[Publisher Full Text](#)
- Thelwall M, Kousha K, Abdoli M, et al.: **Can REF output quality scores be assigned by AI? Experimental evidence.** 2022.  
[Publisher Full Text](#)
- Uddannelses- og Forskningsministeriet: **Aftale mellem regeringen Socialdemokratiet), Dansk Folkeparti, Socialistisk Folkeparti, Radikale Venstre, Enhedslisten, Det Konservative Folkeparti, Nye Borgerlige, Frie Grønne, Liberal Alliance, Alternativet og Kristendemokraterne om: Basismidler til forskning.** 2021.  
[Reference Source](#)
- University of Turku: *Policy for Responsible Assessment of Research and Researcher.* University of Turku; n.d.; 8.  
[Reference Source](#)
- Van Leeuwen T: **Descriptive Versus Evaluative Bibliometrics: Monitoring and Assessing of National R&D Systems.** Moed HF, Glänzel W, Schmoch U, editors. *Handbook of Quantitative Science and Technology Research.* Netherlands: Springer; 2004; (pp. 373–388).  
[Publisher Full Text](#)
- Van Raan AFJ: **Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods.** *Scientometrics.* 2005; **62**(1): 133–143.  
[Publisher Full Text](#)
- Waltman L: **Responsible metrics: One size doesn't fit all.** *CWTS Blog.* 2018.  
[Reference Source](#)
- Waltman L, Kaltenbrunner W, Pinfield S, et al.: **How to improve scientific peer review: Four schools of thought.** *Learned Publishing.* 2023; **36**(3): 334–347.  
[Publisher Full Text](#)
- Watermeyer R, Derrick GE, Borrás Batalla M: **Affective auditing: The emotional weight of the research excellence framework.** *Research Evaluation.* 2023; **31**(4): 498–506.  
[Publisher Full Text](#)
- Wellcome Trust: *What Researchers Think About the Culture They Work In.* Wellcome Trust; 2020; 51.  
[Reference Source](#)
- Wilsdon J, Allen L, Belfiore E, et al.: **The metric tide: Report of the independent review of the role of metrics in research assessment and management.** 2015.  
[Publisher Full Text](#)
- Ylijoki O-H, Lyytinen A, Marttila L: **Different research markets: A disciplinary perspective.** *High. Educ.* 2011; **62**(6): 721–740.  
[Publisher Full Text](#)

# Open Peer Review

Current Peer Review Status:  

---

## Version 2

Reviewer Report 28 May 2024

<https://doi.org/10.5256/f1000research.165965.r279711>

© 2024 E. Hug S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Sven E. Hug** 

Department of Psychology, University of Zurich, Zürich, Switzerland

I'd like to thank the authors for taking the time to consider and address my comments. I have no further comments.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Research evaluation, bibliometrics, peer review

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 27 May 2024

<https://doi.org/10.5256/f1000research.165965.r279712>

© 2024 Pinero de Plaza, PhD D. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Dr Maria Alejandra Pinero de Plaza, PhD** 

College of Nursing and Health Sciences, Flinders University, Adelaide, South Australia, Australia

I have thoroughly reviewed the amended version of the article and have no further comments. I commend the research team for their outstanding work and dedication to this research.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** My research focuses on public health and improving healthcare services. I use

different methods to find effective ways to promote healthy lifestyles and tackle health disparities. I also study how technology affects people's behaviour and how research findings can benefit society. By contributing to evaluation and implementation science, my goal is to turn research discoveries into practical solutions that can improve healthcare delivery, public health outcomes, and the development of strong health and care networks.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

### Version 1

Reviewer Report 29 February 2024

<https://doi.org/10.5256/f1000research.154206.r243163>

© 2024 Pinero de Plaza, PhD D. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Dr Maria Alejandra Pinero de Plaza, PhD** 

College of Nursing and Health Sciences, Flinders University, Adelaide, South Australia, Australia

This work introduces SCOPE, a five-stage framework used to plan, design, and conduct research evaluations. The method has been successfully applied in various organisational settings globally and offers valuable insights for replication by other organisations to adopt similar evaluation practices. Four case studies demonstrate how SCOPE has been utilised to provide value-based research evaluations. The method is part of international efforts towards more meaningful and robust research evaluation practices. The case studies show that SCOPE can be adapted to different organisational settings and levels of detail. The framework's technical soundness is clearly explained in its systematic approach to planning, designing, and conducting research evaluations. The detailed process descriptions and outcomes of each case study further validate the framework's practical application. The approach is used to enhance research culture, develop assessment mechanisms, and design national research assessment exercises.

I suggest enhancing the background section by including insights from seminal works and the latest innovative evaluation methodologies. This addition would provide a robust framing and foundation to introduce the SCOPE framework. Referencing seminal works such as "Evaluation: A Systematic Approach" and "Utilization-Focused Evaluation" can offer comprehensive perspectives to the reader on evaluation methods and stakeholder involvement while incorporating innovative methodologies like "Co-designing, measuring, and optimizing innovations and solutions within complex adaptive health systems," also known as PROLIFERATE. These addendums align with responsible research assessment principles advocated by initiatives like DORA, the Leiden Manifesto and the Hong Kong Principles. Integrating these suggestions into the background would not only contextualise SCOPE's significance within the evolving research evaluation landscape but also demonstrate a comprehensive understanding of the field, thereby

strengthening its contribution to promoting responsible research practices.

### References

1. Pinero de Plaza MA, Yadav L, Kitson A: Co-designing, measuring, and optimizing innovations and solutions within complex adaptive health systems. *Front Health Serv.* 2023; **3**: 1154614 [PubMed Abstract](#) | [Publisher Full Text](#)

**Is the rationale for developing the new method (or application) clearly explained?**

Yes

**Is the description of the method technically sound?**

Yes

**Are sufficient details provided to allow replication of the method development and its use by others?**

Yes

**If any results are presented, are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions about the method and its performance adequately supported by the findings presented in the article?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** My research focuses on public health and improving healthcare services. I use different methods to find effective ways to promote healthy lifestyles and tackle health disparities. I also study how technology affects people's behaviour and how research findings can benefit society. By contributing to evaluation and implementation science, my goal is to turn research discoveries into practical solutions that can improve healthcare delivery, public health outcomes, and the development of strong health and care networks.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Reviewer Report 19 December 2023

<https://doi.org/10.5256/f1000research.154206.r218194>

© 2023 E. Hug S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Sven E. Hug** 

Department of Psychology, University of Zurich, Zürich, Switzerland

The SCOPE framework is a significant milestone in the evolution of the field of research evaluation, as it is the first framework – apart from some societal impact frameworks – that is comprehensive in scope and introduces a procedure for designing evaluations. It is likely that the framework will pave the way for the professionalisation of research evaluation and spark more work on research evaluation frameworks, approaches, and methods.

The three principles and the five steps presented in sections 5 and 6 are sound, consistent, reasonable, and comprehensible. The use cases presented in section 7 are clearly structured and described; they demonstrate the wide applicability of the framework. Even though the three principles and five steps could be specified, modified, and supplemented in many respects, I make no suggestions in this regard for two reasons. First, the framework in its current form is already productive and useful in informing research evaluation practice. Practical experience with the framework will undoubtedly trigger a further development of the framework. Second, the framework can serve as a benchmark and starting point for scholars and practitioners to either supplement the SCOPE framework or develop their own research evaluation frameworks. In my view, it is therefore not necessary nor desirable to change the framework presented in this paper or the accompanying 19-page guide at this time.

While the framework itself is sound, there are three issues with how it is presented in the paper that require attention.

(1) The paper promises too much.

The title promises that the framework will “implement the ideals of responsible research assessment (RRA)”. The paper, however, does not provide a definition of RRA (section 2), only briefly discusses some RRA values and principles (section 3), and describes the relationship between RRA values/principles and the SCOPE framework only briefly and broadly (section 4, table 1).

A minimum solution would be to remove the definite article from the title (implementing the ideals > implementing ideals) and to indicate that there are more RRA values/principles and that RRA values/principles are emerging and therefore vague. A maximum solution would be to provide a definition of RRA (or: a discussion of RRA definitions), discuss and list RRA values/principles comprehensively (or at least in more detail), and describe the relationship between RRA values/principles and the SCOPE framework in detail.

(2) Responsible research assessment is not sufficiently described and discussed.

Section 2 focuses on RRA. The last paragraph suggests that the negative consequences of (all) evaluation methods – and the awareness of them – are a recent phenomenon. Such concerns, however, are almost as old as evaluative bibliometrics and modern peer review. These concerns are therefore not sufficient to explain the surge of reform proposals and initiatives. Moreover, the section does not discuss definitions of RRA or reflect on RRA. This is surprising as one of the authors of this paper has written a blog on this very topic

(<https://www.leidenmadtrics.nl/articles/navigating-responsible-research-assessment-guidelines>).

Section 2 could benefit from a more thorough discussion of RRA, including, for example, Curry et al. (2020), Rushforth and Hammarfelt (2023), Rushforth and de Rijcke (2023), Peruginelli and Pölönen (2023) and Gauffriau (<https://www.leidenmadtrics.nl/articles/navigating-responsible->

research-assessment-guidelines).

(3) The section "Discussion and conclusions" is not very critical.

Section 8 contains a lot of praise for the framework but does not really point out weaknesses and limitations. For example, the authors state that the simplicity of the framework is both a strength and a weakness. In what respect is it a strength and a weakness? What does this imply for practice and the further development of the framework? Moreover, section 8 does not discuss future research.

I therefore encourage the authors to write a more balanced discussion of the strengths, weaknesses, and limitations of the framework, and to delineate future research.

Curiosity question from the reviewer, not to be addressed in the paper:

I count eight examples of use cases in the paper and ten on the webpage, four of which are referenced in the paper. This makes 14 use cases in total. The paper says on p. 10 that "the SCOPE framework is in wide usage by a range of organisations globally." Does this imply that SCOPE is used in more than the 14 cases?

## References

1. S, Curry S, Rijcke A, Hatch D, et al.: The changing role of funders in responsible research assessment: Progress, obstacles and the way ahead dust. *Research on Research Institute*. 2020. [Publisher Full Text](#)
2. G:Peruginelli J;polonen: The legal foundation of responsible research assessment: An overview on European Union and Italy dust. *Research evaluation*. 2023; **00 issue**. [Publisher Full Text](#)
3. A:Rushforth S, Rijcke: Practicing responsible research assessment: Qualitative study of faculty hiring, promotion, and tenure assessments in the United States. *SocArXiv*. 2023; **0** (0). [Publisher Full Text](#)
4. A:Rushforth B, Hammarfelt: The rise of responsible metrics as a professional reform movement: A collective action frames account. dust. *Quantitative Science Studies*. 2023; **1** (37). [Publisher Full Text](#)

**Is the rationale for developing the new method (or application) clearly explained?**

Yes

**Is the description of the method technically sound?**

Yes

**Are sufficient details provided to allow replication of the method development and its use by others?**

Yes

**If any results are presented, are all the source data underlying the results available to ensure full reproducibility?**

No source data required

**Are the conclusions about the method and its performance adequately supported by the findings presented in the article?**



Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Research evaluation, bibliometrics, peer review

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

---

## Comments on this article

### Version 1

Reader Comment 04 Oct 2023

**Lutz Bornmann**, Max Planck Society, Germany

This is an interesting, practical framework for responsible evaluations. In my opinion, the framework can be very well used as guiding schema for evaluations in many contexts. What I miss is the connection to an underlying theoretical perspective. For example, the paper could be stronger rooted in the normative theoretical framework of Robert K. Merton (as it is already). Another possible framework comes from decision sciences (psychology). In a recent paper, Julian Marewski and myself proposed the heuristics framework in this respect (<https://link.springer.com/article/10.1007/s11192-019-03018-x>). Heuristics are simple strategies that, by exploiting the structure of environments, can aid people to make smart decisions in research evaluations.

**Competing Interests:** No competing interests were disclosed.

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

**F1000Research**