

Research assessments should recognize responsible research practices  
— Narrative review of a lively debate and promising developments  
Non Peer-reviewed author version

AUBERT BONN, Noemie & Bouter, Lex (2021) Research assessments should recognize responsible research practices — Narrative review of a lively debate and promising developments.

DOI: [10.31222/osf.io/82rmj](https://doi.org/10.31222/osf.io/82rmj)

Handle: <http://hdl.handle.net/1942/36180>

Title:

# Research assessments should recognize responsible research practices

Narrative review of a lively debate and promising developments

Noémie AUBERT BONN<sup>1</sup> and Lex BOUTER<sup>2</sup>

<sup>1</sup> Amsterdam University Medical Centers, Amsterdam, The Netherlands; Hasselt University, Hasselt, Belgium  
Email: [noemie.aubertbonn@uhasselt.be](mailto:noemie.aubertbonn@uhasselt.be)

<sup>2</sup> Amsterdam University Medical Centers, Amsterdam, The Netherlands; Vrije Universiteit, Amsterdam, The Netherlands

**Submitted as chapter for the book:**

*Handbook of Bioethical Decisions – Vol. II Scientific Integrity and Institutional Ethics*

Edited by Erick Valdés and Juan Alberto Lecaros

# Table of content

---

<b>Abstract</b>	<b>3</b>
<b>1. Brief introduction to research assessments</b>	<b>4</b>
<b>2. Problems and innovative actions</b>	<b>6</b>
2.1 Content	7
2.1.1 An exaggerated focus on research outputs	7
2.1.2 Quantity over quality	10
2.1.3 Inappropriate use of metrics	11
2.1.4 Narrow views of impact	14
2.1.5 Obstacle to diversity	15
2.3 Procedure	17
2.4 Assessors	18
2.5 Research environments	19
2.6 Coordination	21
<b>3. Way forward</b>	<b>22</b>
<b>Acknowledgements</b>	<b>23</b>
<b>Declaration of interests</b>	<b>23</b>
<b>Abbreviations</b>	<b>23</b>
<b>References</b>	<b>25</b>

## Abstract

Research assessments have been under growing scrutiny in the past few years. The way in which researchers are assessed has a tangible impact on decisions and practices in research. Yet, there is an emerging understanding that research assessments as they currently stand might hamper the quality and the integrity of research. In this chapter, we provide a narrative review of the shortcomings of current research assessments and showcase innovative actions that aim to address these. To discuss these shortcomings and actions, we target five different dimensions of research assessment. First, we discuss the content of research assessment, thereby introducing the common indicators used to assess researchers and the way these indicators are being used. Second, we address the procedure of research assessments, describing the resources needed for assessing researchers in an ever-growing research system. Third, we describe the crucial role of assessors in improving research assessments. Fourth, we present the broader environments in which researchers work, explaining that omnipresent competition and employment insecurity also need to be toned down substantially to foster high quality and high integrity research. Finally, we describe the challenge of coordinating individual actions to ensure that the problems of research assessments are addressed tangibly and sustainably.

## 1. Brief introduction to research assessments

Throughout their careers, researchers will face dilemmas and need to make decisions regarding the ethics and the integrity of their work. Earlier chapters in this volume illustrate the substantial challenges and dilemmas involved and the impact that researchers' decisions can have on research, knowledge, and practices. But decisions are not limited to research practices, they also need to be made about researchers themselves. Deciding which researchers should receive grants, which researchers are allowed to start a career in academia, who are promoted, and who obtain tenure are complex issues that shape the way in which research systems operates.

In this chapter, we provide an overview of the complexities of research assessments. More specifically, we provide a critical overview of the problems that current research assessments generate and showcase innovative actions that are introduced with a view to improve the process. We start by briefly introducing research assessments<sup>1</sup> and the debate on whether they are fit for purpose. We then discuss problems of research assessments on five different dimensions: the content; the procedure; the assessors; the environment; and the coordination between these dimensions (Figure 1).

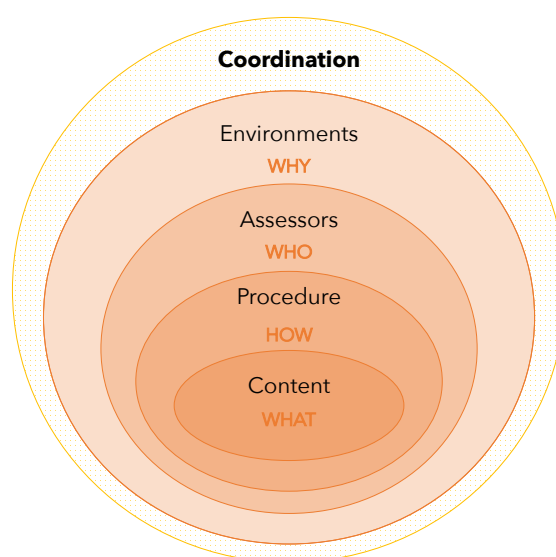


Figure 1. The five dimensions of researcher assessments addressed in this chapter

Research assessments entail important decisions about what matters (i.e., what should be valued in academic careers and research outputs), about who decides what matters, and about how what matters can be measured. In addition to the inherent complexity, the decisions needed for research assessments depend on several stakeholders with their own distinct

<sup>1</sup> Throughout this chapter, we use the term 'Research assessment' interchangeably to refer to the assessment of researchers, research teams, research institutes or research proposals. Given that the term 'research assessment' is most commonly used in current discussions to describe the process through which research resources — be it funding, hiring, recognition, tenure, or promotions — are distributed, we used this term in its broad, interchangeable sense throughout this chapter.

interests. Given the profound complexity, the high stakes, and the many actors involved in such decisions, it is no surprise that research assessments raise substantial controversies. Before introducing the problems and latest innovations in research assessments, we thought that it may help to provide a quick historical snapshot of the evolution of the discourse. This historical snapshot is high-level initially, but we will detail and document each point in greater depth throughout this chapter.

Scientists have scrutinised the attribution of success in academic research for well over half a century [1-3], yet we can pin the beginning of the debate on research assessments on the 1980's, when the growing investments in research led to a substantial growth of the academic workforce [4]. This growth introduced a stronger need for fair distribution of research resources, for example in funding allocation, hiring, tenure, and promotion. Publication metrics which had made their appearance some years earlier — namely publication counts, the H index, citations counts, and journal impact factors — were introduced in research assessments as an opportunity for broad scale, rapid, and comparative research assessment that provides a greater sense of objectivity than traditional peer-review qualitative assessment [5]. Quite rapidly however, it became clear that the newly adopted metrics influenced the publication practices of researchers also in less desirable ways. Early metrics focused on quantity, for instance by using the number of scientific papers researchers published as an indicator of success. This focus on quantity invited high volumes of lower quality scholarly outputs [6]. To address this problem, journal impact factors and citation counts started being used in assessments, asking researchers to place impact before volume. This change had the desired effect and redirected the scholarly output towards prestigious high impact journals [7]. With occasional exceptions, assessors and researchers overall appeared to be satisfied with the new methods until the early 2000's. The beginning of the 21<sup>st</sup> century brought with it a vivid interest in meta-research, research integrity, and bibliometrics. Researchers started understanding that research was vulnerable to misconduct and inaccuracies [8, 9], and that research assessments could influence research in harmful ways [10]. Not only did impact-metrics influence the types of research being done, but they also made research move away from important integrity and quality aspects such as reproducibility and open science [11]. At the same time, the researchers were growing more aware of the high pressure and highly competitive environment they worked in and the impact this had on their work [12, 13]. Consequently, researchers and research communities joined forces to address these challenges and in demanding change in the way in researchers are assessed.

The San Francisco Declaration on Research Assessments [DORA; 14], The Metric Tide [15], and the Leiden Manifesto [16] were among the first key documents to specifically address and raise awareness on the faults of the current assessment. Mostly focused on metrics, these pioneer works were then followed by position statements from numerous groups and organizations who broadened the issue towards research climates, research careers, and research integrity. In Table 1, we showcase a selection of position statements and documents from

general and broad-reaching groups. The eleven documents displayed in Table 1 are only a tiny selection of the booming number of positions papers, initiatives, perspectives, and recommendations now available from different research institutions, research funders, learned associations, and policy groups. Consequently, it would be fair to say that the debate on research assessments has reached strong momentum, and that substantive changes likely are underway.

**Table 1.** Selection of position statements specifically addressing research assessments

Year	Issuing organization	Title
2013	American Society for Cell Biology	San Francisco Declaration on Research Assessments [14]
2013	eLife	Reforming Research Assessments [17]
2013	Science in transition	Why science does not work as it should and what to do about it [18]
2015	Higher Education Funding Council for England	The Metric Tide [15]
2015	Centre for Science and Technology Studies	The Leiden Manifesto [16]
2018	Global Young Academy	Publishing models, assessments, and open science [19]
2018	Moher et al. <sup>a</sup>	Assessing scientists for hiring, promotion, and tenure [11]
2019	European Universities Association	Reflections on University Research Assessments – Key concepts, issues and actors [20]
2020	Science Europe	Position Statement and recommendations on Research Assessment Processes [21]
2020	World Conferences on Research Integrity	The Hong Kong Principles for assessing researchers: Fostering research integrity [22]
2020	Research on Research Institute	The changing role of funders in responsible research assessment: progress, obstacles and the way ahead [23]
2020	Latin American Forum for Research Assessment (FOLEC)	Towards a Transformation of Scientific Research Assessment in Latin America and the Caribbean Series [24-26]
2021	European Open Science Cloud (EOSC) Co-Creation projects	Draft vision for FAIRer assessments [27]

<sup>a</sup>Moher et al. (2018) references several additional papers that address research assessments

## 2. Problems and innovative actions

Changing research assessments is a complex endeavour that requires multiple stakeholders, coordination, and finetuning. In the following sections we introduce a selection of key problems with current research assessments and describe a number of promising actions currently taken to address these problems and improve research assessments.

Problems with research assessments can happen on several interconnected dimensions, some of which are incredibly difficult to tackle. As a starting point, it is essential to address problems with the indicators and the approaches contained in the assessments themselves. But although the *content* of assessments is a necessary starting point for tackling assessments, it is not the only dimension that needs to be addressed to fully make research assessments fit for purpose.

The *procedure* followed and the *assessors* responsible for assessing researchers are also important in enabling changes. Even if the indicators, the procedure, and the assessors are optimal, the research culture plays an additional role in ensuring that changes to research assessments indeed improve the practices and decisions of researchers. Consequently, the *environment* in which researchers work, albeit complex and difficult to address directly, also needs a place in initiatives that aim to change assessments and help foster better research. Finally, a good *coordination* of efforts is needed to ensure that the changes are profound, coherent, and sustainable.

In the following section, we describe key problems and innovative action on the *content*, *procedure*, *assessors*, *environments*, and *coordination* of research assessments. Table 2 summarizes the main points addressed.

## 2.1 Content

Reflection on research assessments should necessarily start with the elements of researcher's professional behavior that are assessed and their impact on the quality and relevance of research. Understanding the problems with the core elements that are used within research assessments is an important starting point to better understand what needs to change.

The problems related to the content of research assessments are too numerous to be able to cover in a book chapter. For simplicity, we selected five key issues that we believe play an important part in the current discourse on research assessments: i) the exaggerated focus on research outputs; ii) the valuation of quantity over quality; iii) the inadequacy of currently used metrics; iv) the narrow definitions of impact; and v) the obstacles current research assessments impose on diversity.

### 2.1.1 An exaggerated focus on research outputs

**The problem.** When looking at research assessments in practice, it is clear that these depend almost exclusively on research *outputs*, most notably on scholarly papers published in international peer-reviewed journals<sup>2</sup>. This focus on outputs has nothing surprising. Considering that a large proportion of research is funded by public investments, it is natural to expect that researchers generate products (in this case research reports) that will ultimately enable tangible benefits for society. Yet, the way in which research outputs are currently measured is problematic in a number of ways.

---

<sup>2</sup> Although research papers are now the most common output currency for career advancement in academia, other indicators such as patents, books, or conference proceedings are also being used in different disciplines. Nevertheless, scholarly papers are dominating the assessment even in disciplines in which they were not common decades ago and in which they have a limited relevance for the transmission of knowledge.



Table 2. Frequent challenges in research assessments and examples of initiatives to improve research assessments

	Problems	Examples of initiatives
CONTENT	<b>Exaggerated emphasis on outputs</b>	<b>Diversify spectrum of indicators</b> Open science badges; Publons, ORCID, open peer review; CRediT; Reporting guidelines (EQUATOR Network) <b>Use assessment models that consider broader activities</b> ACUMEN; OS-CAM
	<b>Quantity over quality</b>	<b>Limit the number of outputs considered</b> Swap full publication lists for a limited number of key accomplishments (e.g., Cancer research UK)
	<b>Inappropriate use of metrics</b>	<b>Raise awareness and mobilize for action</b> DORA; Leiden Manifesto; The Metric Tide; Hong Kong Principles <b>Combine metrics with human input</b> Diverse examples are available in the repository 'Reimagining academic assessment: stories of innovation and change' developed by DORA in collaboration with EUA and SPARC Europe <b>Enable research to find better ways to assess researchers</b> Open Science Policy Platform (e.g., Working Group on Rewards; Expert Group on Indicators; Mutual Learning Exercise on Open Science - Altmetrics and Rewards) <b>Use more comprehensive metrics</b> Altmetrics, PlumX
	<b>Narrow views of impact</b>	<b>Consider a broader spectrum of impact (e.g., societal impact)</b> More comprehensive metrics (see above); RQ+ <b>Allow more open and personal descriptions of impact</b> Narrative CVs and portfolios (e.g., UK Royal Society Resumé for researchers, Health Research Board Ireland; Dutch Research Council; Swiss National Science Foundation)
	<b>Obstacles to diversity</b>	<b>Broaden diversity and inclusion policies</b> Athena Swan; Policies in hiring and promotion; IUPUI recognition of equality, diversity, and inclusion activities in tenure and promotion <b>Ensure greater granularity of research contributions and team dynamics</b> CRediT; DARE <b>Allow more diversity of academic profiles</b> Open University UK diversification of career paths; Ghent University new career track; Dutch Recognition and Reward Programme <b>Enable team recognition</b> Dutch Spinoza and Stevin prizes
PROCEDURE	<b>Assessment time and resource involvement</b>	<b>Reduce the resource involvement needed to review applications</b> Post peer-review lottery (i.e., allocating grants randomly after initial quality check) <b>Reduce the frequency of assessments</b> Longer-term funding; Fewer in-career assessments e.g., Ghent University
ASSESSORS	<b>High potential for biases</b>	<b>Enlarge diversity of assessors' profiles</b> Science Europe recommendations on research assessments; Obtain 360° input from colleagues <b>Avoid biasing elements</b> Avoid adding the applicant's photo to the applications; Move the biography to the end of applications <b>Train assessors to minimize biases</b> Tampere University; HRB Ireland; 'Room for everyone's talent'
	<b>Unclear terminology and undefined abstract concepts</b>	<b>Clearly define the terms used in assessments (e.g., excellence, impact)</b> Dutch Recognition and Reward Programme, Norway Universities
ENVIRONMENTS	<b>Career instability and hyper-competition</b>	<b>Raise awareness</b> ISE Position on precarity of academic careers; Camille Noûs; UK UCU strikes <b>Help researchers have a more secure salary</b> Swedish Regeringskansliet initiative
	<b>Environments not conducive to integrity</b>	<b>Help institutions create healthier research environments</b> SOPs4RI European Commission project
COORDINATION	<b>Lack of coordination and harmonization between stakeholders</b>	<b>Call for more responsible assessment practices</b> European Commission's 'Towards 2030' vision statement; European Commission Open Science Policy Platform <b>Make funding contingent on responsible assessments</b> Wellcome UK <b>Combine efforts</b> Dutch Recognition and Reward Programme; FOLEC; Universities Norway; Responsible Research Network Finland

Note: The initiatives presented in this table are detailed and documented throughout the text. Abbreviations are explained in the Abbreviation section.

For one, the exaggerated emphasis on *research outputs* means that current assessments are oblivious to most of researchers' commitments. Publishing papers, as important as it is, is far from the only activity researchers spend their time and efforts on [28]. Teaching and providing services — the two other pillars of academic careers — and other essential tasks such as mentoring, reviewing or team contributions almost always take second place or are even ignored in research assessments [29]. And within the pillar of 'research', many activities and processes that would provide invaluable information on how the research is conducted are largely ignored from current output-oriented assessments, creating a culture "that cares exclusively about what is achieved and not about how it is achieved" [30]. For example, the detailed methods, the approaches, the specific contributions, or the translation of research in practice are rarely considered in research assessments [31]. This lack of consideration for research processes risks losing sight of important procedural concepts thought to be highly important in advancing science, such as quality, integrity, and transparency [32].

**Innovative action.** In the past few years, there has been an increasing awareness that linking research assessments almost exclusively to research outputs may be problematic [30]. Principle 5 of the Hong Kong Principles, and recommendations 3 and 5 of the DORA directly address this issue, stating that a broader range of research activities should be considered in research assessments. One concrete initiative which may be a first step in solving this problem is the provision of greater visibility to a range of activities that are part of researchers' daily tasks. The Open Science badges — registration, open data, open materials — are a good example of a simple change that allows readers or eventually assessors to quickly capture open science practices behind published works [33]. The presence of reporting guidelines, such as those available on the EQUATOR (Enhancing the QUALity and Transparency Of health Research) Network [34] can also summarize details and procedures and provide information on the transparency and reproducibility of the work. The increasing availability of open and transparent peer-review and initiatives that provide visibility of peer-review commitments such as Publons [35] or ORCID (Open Researcher and Contributor ID) [36] are other examples that can help enrich the indicators used to assess researchers. The Contributor Role Taxonomy (CRediT) which provides more information on the roles, and responsibilities that researchers take is another example we will discuss further in section 2.1.5 [37, 38].

Broader indicators are increasingly visible in more formal assessments procedures. For instance, the Academic Careers Understood through MEasurement and Norms (ACUMEN) portfolio provides a template that considers indicators from a very diverse array of activities [39]. While the ACUMEN remains largely quantitative, its broad coverage of research activities is a good reminder that assessments can be much more comprehensive. The European Commission's Open Science Career Assessment Matrix (OS-CAM) is a similar model of assessment that includes a broad array of research activities such as teaching, supervision and

mentoring, professional experience and even has an explicit section on research processes [40]. We will discuss other ways of broadening assessments such as narrative CVs and portfolios in section 2.1.4.

### 2.1.2 Quantity over quality

**The problem.** Another important problem of researcher assessments is their tendency to value quantity over quality. Many researchers feel encouraged to publish as many papers as possible and are sometimes offered tangible incentives such as financial rewards to publish more [41, 42]. Assessing researchers on the number of published papers does indeed lead to more publications, but it tends to do so at the detriment of research quality [6, 43]. It can also encourage questionable research practices such as 'salami slicing' — "the spreading of study results over more papers than necessary" [44] — and can tempt researchers to favour journals where acceptance rates are high rather than journals suited for their work or journals with thorough peer-review procedures. Unsurprisingly, the longing for quantity also works in favour of predatory publishers and paper mills whose business model is targeting authors desperate to publish regardless of quality [45, 46].

To address this problem, research and funding institutions are increasingly modifying their assessment procedures to focus on impact rather than on quantity. Nevertheless, the impressive numbers of peer-reviewed publications or books that are very often stated in researchers' biographies reminds us that productivity is still considered an important indicator of accomplishment within the research community and the research culture. Quantity indicators also remain key to institution-level assessments, a point we will discuss further in the Coordination section.

**Innovative action.** The obvious solution to reduce the focus on quantity should be to look more at quality. But even though ways to assess quality are starting to pierce, the endeavour is a bit more complex than it may seem. For example, Eyre Walker and colleagues showed that, when scientists assess a published paper without knowing the journals in which the paper was published, they are generally inconsistent and unable to judge its intrinsic merit or to estimate the impact factor of the journal in which the paper was published [47]. However, assessing quality of publications is not the only way assessments can deviate from quantity indicators. In the past few years, several research and funding institutions diverted assessments away from quantity by asking researchers to select only a subset of their work — generally three to five key accomplishments or contributions (e.g., publications, events, changes in practice, committee participation, etc.) — and to describe why these accomplishments matter (see for example [48]). Focusing on a limited number of outputs enables a more in depth assessment which is likely to refocus the assessors' attention away from quantity towards content, meaning, and quality.

### 2.1.3 Inappropriate use of metrics

**The problem.** As we mentioned above, most research assessments swapped volume-metrics for impact-metrics to incite researchers to publish in more prestigious journals. Among those, the journal impact factor, citations counts, and the H-index raise important challenges.

Of all impact-informed metrics available, the journal impact factor is probably the most widely used in current research assessments. In a review of their use in North American academic review, promotion, and tenure document, McKiernan and colleagues found that 40% of research intensive institutions explicitly mention journal impact factors [49]. The journal impact factor of a given year is the ratio between the number of citations received in that year for publications in that journal that were published in the two preceding years and the total number of "citable items" published in that journal during the two preceding years. [50, 51]. The journal impact factor was designed to help librarians select the journals they should subscribe to, but it was never intended to influence researcher evaluations. On the contrary, Eugene Garfield — widely known as the father of journal impact factors — explicitly warned against using journal impact factors for assessing individual scholarly articles [52]. Nevertheless, the seductive power of a single metric that would allow to quantify the 'value' of journal articles quickly won over research assessments. Unfortunately impact factors introduced substantial problems of their own. First, the mere fact that journal impact factors became recognized as a measure of success reduced their objectivity as a measure of success; a phenomenon known as Campbell's law [53]. In fact, journal impact factors incite strategic responses from researchers, many of which are now considered to be questionable research practices. These include among others selective reporting, 'spin', p-hacking, HARK-ing (hypothesizing after results are known) and non-publication of negative results [5, 7, 54, 55]. Journal impact factors further suffer from fundamental weaknesses that allow them to be gamed relatively easily [56].<sup>3</sup> In addition, impact factors are a journal-level metric and are therefore not a valid measure for the impact of individual papers or of the authors of that paper. Indeed, the distribution of citations in a journal tends to be so skewed that impact factors provide little information on the number of citations individual papers in that journal can expect [57, 58]. Finally, by the way journal impact factors are calculated, they ignore slow citation (i.e., citations two or more years after publication), thereby potentially bias against innovative research [59]. Despite these fundamental flaws, journal impact factors are still widely used in researcher assessments and are frequently described as an indicator of the quality of individual research papers [31].

Without even entering the colossal debate on the relationship between citation metrics and research quality, it may be relevant to consider the actual number of citations which are also frequently used in researcher assessments despite the fact that these require more time to

---

<sup>3</sup> From these problems, we can mention the unequal citation practices for different topics or article types as well as the imbalance between the numerator — which contains all citations to a journal for the given years — and the denominator — which only contains the number of 'citable items', and thereby excludes editorials, commentaries, news and views, and other items that are increasingly taking predominance in high impact factor journals [7, 56].

accumulate. Citations are problematic in different yet connected ways. To begin, numbers of citations provide no information on the reasons a paper is cited. Citations used to provide background information, to build an argument, to support a theory, to raise a problem, or to criticize a paper all count in the same way [51]. Citations can also be manipulated, for example through peer-reviewer or editor requests, or by forming citation cartels [60, 61]. They are also prone to biases unrelated to the intrinsic merit of a paper [62]. And finally, direct citations are often only partially and sometimes not at all supported by the cited article, suggesting that researchers often cite papers without reading or even downloading them [63].

The H-index — or Hirsch Index for its inventor Jorge E. Hirsch — is another indicator that is frequently used in research assessments. The calculation is quite simple: a researcher has an h-index of  $x$  when she or he published at least  $x$  papers which were cited at least  $x$  times each. In other words, the h-index combines impact and productivity to provide information at an individual level. Nonetheless, the H-index is also strongly criticized. First, the misleading simplicity of a single number to judge researchers is already problematic, especially when comparing researchers from different fields of expertise. Furthermore, although the H-index combines paper and citation counts, it will never be higher than the total number of papers a researcher has published, regardless of the number of citations these papers have (e.g., a researcher with 10 papers cited 10 times each will have a higher H-index than a researcher with 9 papers cited 100 times each) [51]. Similarly, as an ever-growing metric, the H-index provides senior researchers with a clear advantage that makes them largely invincible when compared to junior researchers, even after they stop being active in research. Jorge E. Hirsch himself stated that the H-index could “fail spectacularly and have severe unintended negative consequences” [64, p.4], and several metrics experts have deemed it inappropriate in measuring researcher’s overall impact [65]. Despite all this, the H-index continues to be used often in research assessments.

Although many other metrics exist, the journal impact factor, citation count, and H-index are the three most frequently used in researcher assessment. On top of their individual flaws, an overarching criticism of these metrics is that they fail to capture the core qualities they aim to measure. More specifically, while several institutions use these metrics as a proxy to assess the quality and impact of the work [49], they provide very little information that could be validly interpreted as *quality* or *impact* [31]. Instead, these metrics provide information on the visibility, the attention, and the citation patterns within academia [51, 66]. Garfield himself qualified citations as an indicator of “the utility and interest the rest of the scientific community finds in [the work]” [67, p. 372], not as a measure of quality. Knowing that impact-informed metrics are even believed to “discourage rigorous procedures, strict replication/confirmation studies and publication of negative, nonstatistically significant results”, it is important to rethink how we use — or at least interpret — impact metrics [68].

Once again however, reinterpreting the role of impact metrics on research assessments requires changes at the core of research communities. Researchers who have spent decades

building a career on inadequate indicators may find it daunting to give up their high rankings to adopt a new system in which they may rank less excellent or even poorly. Increased awareness, discussion, and mobilisation are still needed.

**Innovative action.** The Declaration on Research assessments [DORA; 14] strongly advocates against using the impact factor in individual research evaluations<sup>4</sup>, supports the consideration of value and impact of all research outputs, and argues that evaluations of scientific productivity must be transparent and explicit. Along the same line, the Leiden Manifesto and The Metrics Tide pledge for the development and adoption of better, fairer, more transparent and more responsible metrics [15, 16]. These three initiatives, recently joined by the Hong Kong Principles for assessing researchers [22], play a crucial role in raising awareness about the shortcomings of widely used research metrics. Awareness is only the first step towards actual change but these initiatives have brought together a community that supports the change. DORA already has nearly 20,000 signatories — over 2000 of which are organizations. And changes are indeed starting to happen at the research institutions, funders, and policy level. For instance, several research institutions now make sure that metrics are not used in isolation, but only as a complement to reflective qualitative peer-review (examples of institutions that have concretized these changes are available in the repository '*Reimagining academic assessment: stories of innovation and change*' developed by DORA in collaboration with EUA and SPARC Europe [69]).

As part of the Horizon 2020 program, the European Commission also created an Open Science Policy Platform in which several expert groups were created to discuss better research assessments and indicators. These include the Working Group on Rewards, the Expert Group on Indicators, and the Mutual Learning Exercise on Open Science – Altmetrics and Rewards [70].

New metrics are also becoming available to help balance research assessments. Simple paper downloads, for example, may capture readers who do not cite works, such as non-academic users of the work [71]. More complex composite metrics have also been built. Altmetrics is a prime example of the diversification of the elements that can be captured on a single piece of work. Altmetrics include a wide array of inputs, such as open peer reviews reports, social media capture, citations on Wikipedia and in public policy documents, mentions on research blogs, mass media coverage, and many more aspects which help provide a broader overview of how the work is being used. The PlumX metrics, although governed by different calculations, works in similar ways. These innovative metrics are gaining increasing visibility on publisher's websites, but their use in formal researcher assessment is still very limited.

---

<sup>4</sup> In fact, DORA's first principle states directly that assessors should "not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions" [14].

### 2.1.4 Narrow views of impact

**The problem.** In addition to the overreliance on outputs and the problem of inadequate metrics we delineated above, indicators currently used in research assessments can be criticized because they provide a very narrow view of research impact. Two main dimensions deserve to be discussed here.

The first dimension concerns the impact research has on practice, policies, or society. As we previously mentioned, researchers are often expected to dedicate a portion of their time to the key pillar of 'Services', but typically their involvement in 'Services' is almost entirely absent from researcher assessments [29]. In addition, in the rare instances where 'Services' are considered in review, promotion, and tenure assessments, their consideration almost exclusively targets services provided within the institution or the research community — such as participation on university boards or editorial boards — rather than services provided to the public or to society [72]. Citations-based metrics only consider recognition and visibility within the scientific (and citing) community and provide only a restricted view of academic impact [73]. Impact on practice, policy and society are not captured and are even obscured by these narrow metrics. For example, the need to publish in high impact factor journals often translates in a need to publish in English-language international journals; a decision that can reduce the societal impact of locally relevant research projects [74]. Academic environments themselves, through their funding objectives, missions, and expectations, value discovery but largely disregard how we can best implement discoveries in practice [75].

A second dimension that is important to reconsider is the impact that research has on knowledge advancement. In fact, current assessments tend to conflate impact with ground-breaking findings [31]. While this idea has long been embedded in the notion of scientific discovery, it also undermines the importance of non-ground-breaking work in advancing knowledge. Borrowing the words of Ottoline Leyser, chief executive officer of UK Research and Innovation:

*"It is worth remembering that the term "ground-breaking" comes from construction. There is often a ground-breaking ceremony, but then the building must be erected. This comes only after much preparation, from determining the ideal location to securing all the planning permissions. Likewise, for every ground-breaking discovery, a huge amount of work has paved the way, and follow-up work to solidify the evidence and demonstrate reproducibility and generality is essential. High-quality work of this sort is rarely recognized as excellent by the scientific enterprise but is excellent nonetheless, and without it, there would be no progress." p. 886 [76]*

The overemphasis on ground-breaking discovery has shaped a research system in which replication studies and negative results are largely invisible despite their crucial value in solidifying knowledge [77-79].

**Innovative action.** To better capture the impact that research has on practice, policies, society, or research itself, research assessors need to broaden the scope of indicators they use. We already mentioned that alternative metrics can help capture interest that would otherwise be missed. Another notable effort that may help capture societal impact in research is the Research

Quality Plus (RQ<sub>+</sub>) evaluation approach used at the International Development Research Centre (IDRC) in Canada [80]. Although emphasising expected impact in a funding application is sometimes criticized for being artificial and highly theoretical [81, 82], the RQ<sub>+</sub> provides a structured method through which societal impact can be estimated before the research takes place. Since the RQ<sub>+</sub> is used for evaluating research proposals, it is not directly applicable to assessing researchers' past accomplishments. Nonetheless, it might be a good model to inspire areas of impact that could be considered in future research assessments.

To capture the impact that the research has in building knowledge, several research institutions and funders started adopting narrative CVs in which researchers are encouraged to describe, in their own words, the impact of their work. A good example of these narrative CVs is the *Résumé* for researchers provided by the Royal Society in the UK [83]. In the *Résumé* for researchers, applicants are provided with unstructured space to discuss their contributions to the generation of knowledge, the development of individuals, the wider research community, and the broader society. These open descriptions enable assessors to consider a broader, more diverse, and more personal perspective of impact that may have been invisible otherwise. While these narrative CVs are not easy to write and more demanding to assess than quantitative metrics, they are increasingly adopted in research institutions. Several other funders, such as the Health Research Board Ireland, the Dutch Research Council, and the Swiss National Science Foundation are also experimenting with open and narrative CVs [84].

### 2.1.5 Obstacle to diversity

**The problem.** In addition to the issues presented above, current research assessments also often fail to promote diversity and inclusion in research. Gender inequalities, for example, are seen in both citation metrics and publication outputs [85, 86], even more so in the disrupted working conditions of the COVID-19 pandemic [87, 88]. Women are also more likely to be strongly involved in teaching, in the hands-on facets of research, or in other contributions that are essential to science but are less likely to result in first- or last-author publications [89, 90]. Similar issues also afflict ethnic groups and geographic regions, not only in funding opportunities and access [91], but also in the fair attribution and recognition of their work [92, 93]. The same hurdles are faced by researchers with disability, even when policies are in place to tackle the injustice [94]. Consequently, research assessment's excessive reliance on publication metrics may further tax diversity and inclusion issues in academia. But diversity and inclusion is not only about disadvantaged groups. Diversity of skills, contribution, and career profiles is also an essential aspect that is largely ignored in current assessments and inclusion policies. Indeed, research assessments tend to assess researchers individually and to expect them to fit a one-size-fits-all model of success in research [31]. This individual and uniform model of assessment contradicts the highly collaborative, differentiated, and complementary roles that are intrinsic to research [95]. Overlooking the still growing



differentiation of research tasks disregards the unique contributions from non-leading members of research teams as well as the essential role of research support staff [96]. Individual assessments and uniform expectations also increase competition between researchers; a feature which is known to be highly problematic and is often mentioned as a cause for research misconduct and questionable research practices [12, 97].

**Innovative action.** The lack of diversity in research is a priority on the agenda of several large funders and research organisations. The Athena Swan Charter, for example, plays an important role in inciting research institutions to achieve gender inclusivity [98]. Several institutions already have internal policies, quotas, and initiatives to promote greater diversity in hiring and promotion, yet some of these policies have raised hefty debates in the past [99, 100]. Going one step further, the Indiana University – Purdue University Indianapolis (IUPIU) decided not only to encourage activities that promote equality, diversity, and inclusion, but also to recognize their inherent value by considering them in researchers' tenure and promotion application [101]. Despite these important initiatives, the impact that the indicators used in assessing researchers have on diversity and inclusion is rarely addressed, and there is growing realization that diversity and inclusion should be more prominent in research assessments [102].

The role an individual has in the research team has also received increasing attention in the past few years. Assessors realise that knowing the ways in which researchers collaborate can provide invaluable information. As a result, interesting initiatives that enable greater visibility on the team aspect of research are starting to pierce. The Contributor Role Taxonomy (CRedit), for example, provides an added level of granularity to authorship and helps to understand the dynamics, roles, and responsibilities in team research [37, 38]. Although contributor roles have not yet fully secured their place in research assessments, more and more journals provide contributorship sections to the papers they publish. Whether the future of academia is one in which contributor roles take over authorship, however, remains to be seen [103, 104]. Another interesting initiative in the recognition of teamwork is the Diversity Approach to Research Evaluation (DARE) [105]. The DARE approach provides tools to measure and understand how collaborators connect and deal with diversity. While the approach is more informative than evaluative, knowing more about the dynamics in research teams is a starting point to gather information on the characteristics of strong research teams.

There is also a growing belief that the lack of diversity in the profiles of individuals that succeed in academia may weaken effective team work [106]. Diversifying the profiles of academic employment, therefore, may help build research climates in which success comes from joint efforts rather than from competition between individuals. One early example of such initiative is the Open University in the UK, where more flexibility is given to researchers to enable to focus on different pillars of their work [107]. As a result, researchers could pursue a career in which knowledge exchange is valued before their teaching and research achievements. The recently implemented career track at Ghent University, Belgium and the Dutch

Recognition and Reward Programme are two other well-known initiatives to address the need for diversifying researchers' profiles [108, 109]. The position paper 'Room for everyone's talent' from the Dutch Recognition and Reward Programme nicely illustrates how such a diversification may take shape. Specifically, researchers have the opportunity to select a unique combination of key areas they wish to specialise in and be assessed on. These key areas include research, education, impact, leadership, and patient care. While all researchers are expected to demonstrate sufficient competencies in the research and education areas, they can choose the extent to which they favour these and any other areas and can change areas of specialties at different stages of their career.

Finally, the initiative contains a clear acknowledgement of the need to reward team efforts. The Dutch's highest research awards, the Spinoza and the Stevin prizes, are now also open to team applications, making another step forward in the recognition of research as team work [110].

## 2.3 Procedure

**The problem.** Changing researcher assessments is a complex endeavour that extends far beyond the elements and indicators assessed. It is also important to discuss the time and resource commitments that research assessments imply.

Researchers need to invest substantial time in building a prestigious CV and in applying for research funding. While the peer-reviewed process through which research is funded is most likely essential for good quality research, the low success rate of current funding schemes (typically 5-10% of the applications are granted) suggests that a lot of efforts are ultimately wasted. Past research has shown that many researchers consider the preparation of funding proposals to be the most *“unnecessarily time-consuming and ultimately most wasteful aspect of research-related workload”* p. 41 [111] and that researchers wished they could spend less of their time on it [112]. In fact, Herbert and colleagues estimated that the amount of time spent preparing grants for the Australian National Health and Medical Research Council in 2012 [113] reached 550 working years of researchers' time — the equivalent of 66 million Australian dollars (around 42.5 million Euros at the time of writing). Considering the low success rate of these applications, competitive funding channels come with phenomenal research time investments. Building a tenure dossier and applying for different research positions is also no small task, and since grants and non-tenured research positions are typically short-term, the time investment involved is substantial.

In turn, the colossal demands for research money and opportunities also lead to increasing numbers of applications which raise faster than the investments in research funding [114]. This growing demand creates a pressure on funders who face an excess of applications to review, and who will, in turn, require peer reviewers and selection committee members — most of the time

researchers themselves — to invest their already scarce research time in the review process [5, 115].

**Innovative action.** With the large amount of demands for funding and career opportunities, it is difficult to reduce the volume of research assessments. Nevertheless, there are ways in which the time and resource investment can be reduced to alleviate the burden of both researchers and assessors. One such initiative is the post-peer-review lottery of funding applications which proposes that, after a first thorough quality check to select proposals that are sound and methodologically adequate, assessors should select the winning applications randomly rather than through lengthy deliberation. This radical idea would not only increase efficiency of research funding assessments [116], but it would also guard against the 'natural selection of bad science' by allowing unusual and unfashionable topics with high risk of negative findings to be funded [117]. The lottery approach may even help reduce career insecurity in academia, a point we will discuss further in section 2.5 [118]. Another way to reduce the burden of research assessment is to reduce the frequency at which researchers are evaluated. Longer terms funding and research contracts could help in this matter, while further alleviating worries around the lack of security of research careers. Similarly, reduced evaluative frequency for employed researchers may help reduce the evaluative burden. Ghent University is currently experimenting this change in its new career track, moving from a review interval of 3 year to one of 5 years starting in 2020 [108].

## 2.4 Assessors

**The problem.** The assessors themselves are not so frequently on the agenda for change to research assessments, despite their direct relevance to assessment processes. Particularly, when reflective and qualitative peer-review takes precedence, a great deal of subjectivity is introduced in the assessment process. Subjectivity is not a bad thing but it leaves substantial room for personal biases and involuntary discrimination in research assessments. For instance, assessors will naturally be tempted to cherry pick the information that confirms their already formed opinion (confirmation bias), to base their assessment on easily accessible anecdotal information (accessibility bias) or to let contextual aspects such as the reputation of universities listed on the CV of applications shape their views of individual candidates (halo effect [see for e.g., 119, 120]), to name only a few [53]. In addition, many assessment procedures ask assessors to value highly abstract concepts — for example 'excellence', 'high impact' — differences in interpretation, misunderstandings, and unfortunately biases can then easily happen [121].

**Innovative action.** Diversity is an important keyword if we want to reduce the influence of biases. Indeed, guidelines explicitly recommend that research and funding organisations should

strive to ensure that reviewer pools and hiring committees contain diverse profiles [21]. In addition, diversity should target not only gender and ethnicity, but also the profiles of assessors and their seniority. For example, there is increasing realisation that the input for researcher assessments, for example the reference letters used, should come from superiors as well as from those supervised or managed by the researcher being evaluated [i.e., 360° feedback; 122]. Other ways to reduce biases on research assessments have been proposed, for example avoiding photos of the candidate on the application or moving educational history with potentially biasing university names to the end of the evaluation, but the efficacy of such approaches remains largely undocumented [123]. Finally, training assessors to ensure that they have a clear understanding of the assessment process and providing unambiguous definition of the key concepts that are assessed (e.g., impact, excellence, quality, etc.) can help reduce biases [21, 121]. A few universities and organisations are starting to implement these recommendations. For example, Tampere University now informs and trains evaluators across campus about responsible evaluation practices [69]. Similarly, the Health Research Board (HRB) Ireland also started raising awareness, training staff, and providing guidance for reviewers as a way to minimize gender inequalities and reduce unconscious biases [124], much like the Dutch Recognition and Reward Program in which training and instructions are provided to assessment committees [109]. Others also started defining the terms they use to assess researchers. For instance, Norway Universities added clear definitions of the key concepts needed in assessments [69], while the 'Room for everyone's talent' position paper explicitly defines the concept of impact. Such initiatives are still scarcely exploited and not yet evaluated, but there is growing awareness of the need to inform, train, and support those who assess researchers.

## 2.5 Research environments

**The problem.** We know that the environments in which researchers operate are problematic since they impose high pressures on researchers to perform and publish [125-127]. Changing research assessments can likely help to reduce the 'publish or perish' culture. Yet, other elements in the environment of researchers are also important to consider to avoid wasting the huge efforts invested in changing research assessments.

First, the lack of stability in research careers is an essential aspect to consider. At the moment, there is a huge discrepancy between junior (temporary) and senior (permanent) positions in academia, and only between 3% and 20% (depending on the countries' estimates and faculties) of young researchers will be able to pursue the career in academia to which they aspire [128-134]. In turn, this lack of stability creates an unhealthy working environment in which stress, mental health issues, and burn out thrive [135-137]. Furthermore, the scarcity of senior positions creates a perverse hyper-competition between junior scientists who wish to

survive in academia. Hyper-competition not only worsens the situation, but it is also known to be an important driver of questionable research practices [12, 97].

Beyond these interpersonal issues, the support, resources, and infrastructures that researchers receive is also essential to ensure that changes in research assessments are implemented effectively. Currently, junior researchers and PhD students often feel unsupported [138, 139] and the transition towards new expectations can generate frustration if the resources to fulfil these new expectations are lacking. For example, expecting researchers to preregister their research protocols or to make their data open and FAIR (i.e., Findable, Accessible, Interoperable, and Reusable [140]) is a great step towards better research, but it comes with important needs for adequate infrastructures, training, and most importantly researchers' time. Similarly, demanding open access publication is increasingly requested by funders and institutions, but it needs to come with a budget for covering article processing charges, without which inequalities may ensue [32].

**Innovative action.** There are several initiatives that aim to improve research environments, and in many ways, the innovative actions mentioned throughout this chapter would help create a healthier, more collaborative research climate. Yet, we would like to provide more details on a three types of initiatives that target research environments directly. First, there are initiatives that play a crucial role in raising awareness and opening the discussion on the problem. Examples include the Initiative for Science in Europe (ISE) position paper on precarity in academic careers and its associated webinar series [118], the French movement of 'Camille Noûs' from Cogitamus Laboratories [141], and the University College Union strikes that took place at 74 Universities across the UK in early 2020 to denounce — among other things — the casualization and the lack of employment security of research careers [142]. Second, more forceful initiatives also start to appear. For instance, at the end of 2020, Sweden produced a national bill to change to the way in which it funds research so that a greater share of researchers' salary would come from governmental non-competitive funding [143]. This bill came in response to a thorough investigation in which it was discovered that the constant search for competitive funding ultimately undermined research quality [144, 145]. In helping researchers to have a more stable salary, Sweden aims to reduce the hyper-competition and to lower the employment insecurity of researchers. The third initiative that is highly relevant when discussing research environments is the Standard Operating Procedures for Research Integrity (SOPs4RI) European Commission project that is ongoing until 2022 [146]. The SOPs4RI project is creating a toolbox of best practices and guidelines to help research and funding institutions build research integrity promotion plans. In doing so, the SOPs4RI emphasizes that research integrity is not only a responsibility of researchers, but also of research and funding institutions whose operating procedures should foster healthy research environments. Simultaneously, the project is also empirically creating its own guidelines on topics that are overlooked in existing research guidance documents. One of the guidelines being produced

directly targets ways in which institutions can build better and more collaborative research environments that foster research integrity.

## 2.6 Coordination

**The problem.** The final point that we find important to discuss is the need for thorough, intense, and continued coordination between different actors of the research system. In fact, to fully address the problems we described in this chapter, an open dialogue and thorough coordination between researchers, funders, research institutions, and policy makers as well as other actors such as publishers and metrics providers is needed.

Without coordination between stakeholders, changing research assessments is difficult and unlikely to happen on a large scale. For instance, in many countries, governments use performance-based attribution to fund research institutions, meaning that the share of funding received by research institutions largely depends on quantity indicators of outputs [147]. Although using bibliometric indicators to distribute funding at an institutional level does not mean that universities should assess researchers using the same criteria [148], the fear of underperforming often leads universities to use these indicators internally at a researcher-level [106, 149]. Similarly, the way in which universities are recognized is profoundly influenced by university rankings. University rankings strongly depend on impact factors and other publication metrics, and there is increasing awareness that they have profound flaws and should be interpreted carefully [150]. Yet, rankings are still a dominant way of attracting funding, researchers, and students, and most universities take strategic, organizational, or managerial action to improve their rankings [151]. Lack of coordination with metrics-providers also play a role in the problem. In fact, most major metrics belong to profitable companies whose external agendas differ from those of the research communities [152]. Thorough communication with publishers is needed if we hope to shape metrics that align with the objectives of the research communities.

Changing researcher assessments is also something that is difficult to implement in single institutions. In the absence of a common approach of research assessments, there is a worry that researchers building a profile to succeed in one proactive institution may later be penalised if they want to migrate to another research setting in which their profile might be undervalued. In other words, the perceived ‘first-mover’s disadvantage’ favours a stagnant status quo and builds a feeling of hopelessness that the highly needed changes will occur [106].

**Innovative action.** Ensuring the coordination of all stakeholders around the same objectives — and finding the means to achieve these objectives — is an extremely challenging task. Among others, the European University Association (EUA) briefing and The Metric Tide provide insights on this crucial need for coordinating actions at the level of research assessments, not hiding the complexity of the tasks it implies [15, 20]. Despite the challenge, best practice

examples mentioned throughout this chapter have shown that coordinated changes are possible in practice.

Actors with broad influence and substantial budgets are essential here. For example, the European Commission's 'Towards 2030' vision statement addresses the issue of ranking, calling research institutions to move beyond current ranking systems for assessing university performance because they are limited and "overly simplistic". [150]. Broad reaching groups such as the European Commission Open Science Policy Platform we mentioned earlier and DORA also plays a role in coordinating changes by uniting different research institutes and member states to agree on a strategic plan of action. In South America, the Latin American Forum for Research Assessment (FOLEC) provides a platform for discussion between stakeholders on issues of research assessments [24-26]. University alliances can also help coordinate changes. For example, in 2019 the consortium Universities Norway put together a working groups aiming to build a national framework for research career assessments. The group issued a report in 2021 in which they propose a toolbox for recognition and rewards of academic careers [153]. The Academy of Finland went through a similar process to create national recommendations for responsible research evaluation [154], and more and more university associations and academies are following this lead.

In a slightly more drastic approach, since 2021 the major UK research funder Wellcome decided that it would only provide funding to researchers working in organizations that can demonstrate that their researcher assessments are fair and responsible [150]. This strategic decision incites efforts from both the institution, which would be at a disadvantage if it did not work to ensure its eligibility to Wellcome funding, and the researchers who will push their institutions to ensure they remain eligible for this important source of funding.

Finally, the program 'Room for everyone's talent' we described above is an inspiring example to prove that profound coordination is possible. In 'Room for everyone's talent', five public knowledge institutions and research funders joined forces to ensure that Dutch research institutions would abide by the new assessment models. In addition, in the position paper announcing the new model, the five parties acknowledge their responsibility to take steps towards even tighter coordination. The position paper describes their commitment to connect with international organisations such as the European University Association, Science Europe, and Horizon Europe to encourage changes and harmonisation at a European level.

### **3. Way forward**

Changing researcher assessments is difficult and requires huge investments and efforts from a diverse array of stakeholders. We have argued that current research assessments have profound inadequacies, but that promising pioneering actions are starting to address these inadequacies and to align research assessments with responsible research practices.

To continue moving forward, we need to think of research assessments in their entire complexity, addressing not only their content, but also the processes, assessors, environment, and coordination needed for change. For each dimension, we must understand the problem, raise awareness, take action, and coordinate efforts to enable change.

Even though research institutions, research funders, and policy makers have a clear responsibility in enabling the change towards more responsible assessments, we, as researchers, also have an important role to play. For one, we should remember the biases and problems of research assessments when acting as peer-reviewers or assessors and ensure that we avoid shortcuts and biases as much as we can. But we should also play a role in shaping the tenacious research culture, helping to raise awareness and mobilise action around us. In the end, when we look at what was accomplished by DORA — which started from a small group of researchers and editors within the research community — researchers can help to drive the change.

But changing research assessments is not the end in itself. To avoid falling in the same pitfalls we are fighting with today, it is essential to understand whether the changes to research assessments help contribute high quality and high integrity research [11]. In this regard, research on research assessments is essentially important to allow us to understand, inform, and realign research assessments towards a better future. In short, we need evidence-based research assessment policies.

## **Acknowledgements**

The authors would like to thank Anna Hatch whose invaluable feedback has helped us improve and enrich this chapter.

## **Declaration of interests**

The authors are involved in several initiatives that aim to improve research assessments. From the initiatives mentioned in this chapter, NAB is post-doctoral researcher in the SOPs4RI consortium and LB is a member of that consortium as well and was an author of the Hong Kong Principles and is active in disseminating them.

## **Abbreviations**

ACUMEN: Academic Careers Understood through MEasurement and Norms

CRedit: Contributor Role Taxonomy

DARE: Diversity Approach to Research Evaluation

DORA: San Francisco Declaration on Research Assessments



EOSC: European Open Science Cloud  
EQUATOR: Enhancing the QUALity and Transparency Of health Research  
EUA: European University Association  
FAIR: Findable, Accessible, Interoperable, and Reusable  
FOLEC: Latin American Forum for Research Assessment  
HRB: Health Research Board Ireland  
IDRC: International Development Research Centre  
ISE: Initiative for Science in Europe  
IUPIU: Indiana University – Purdue University Indianapolis  
KNAW: Royal Netherlands Academy of Arts and Sciences  
NFU: Netherlands Federation of University Medical Centres  
NOW: Dutch Research Council & Institutes  
ORCID: Open Researcher and Contributor ID  
OS-CAM: Open Science Career Assessment Matrix  
RQ+: Research Quality Plus  
SOPs4RI: Standard Operating Procedures for Research Integrity  
UCU: University College Union  
VSNU: Association of Universities in the Netherlands

## References

1. Merton RK. Priorities in Scientific Discovery. In: Storer NW, editor. *The Sociology of science : theoretical and empirical investigations* (1973). Chicago (IL): University of Chicago Press; 1957.
2. Zuckerman H, Merton RK. Patterns of Evaluation in Science: Institutionalisation, Structure and Functions of the Referee System. *Minerva*. 1971;9(1):66-100.
3. Hagstrom WO. Competition for recognition. *The scientific community*. Carbondale (IL): Southern Illinois University Press; 1975. p. 69-104.
4. Alberts B, Kirschner MW, Tilghman S, Varmus H. Opinion: Addressing systemic problems in the biomedical research enterprise. *Proceedings of the National Academy of Sciences*. 2015;112(7):1912.
5. Gingras Y. *Bibliometrics and Research Evaluation: Use and Abuses*. Cambridge (MA): MIT Press; 2016. 136 p.
6. Butler L. Modifying publication practices in response to funding formulas. *Research Evaluation*. 2003;12(1):39-46.
7. Larivière V, Sugimoto CR. The Journal Impact Factor: A brief history, critique, and discussion of adverse effects. *arxiv*. 2018. doi:arXiv:1801.08992v2
8. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2(8):696-201.
9. Martinson BC, Anderson MS, De Vries R. Scientists behaving badly. *Nature*. 2005;435(7043):737-8.
10. Abbasi K. Let's dump impact factors. *Br Med J*. 2004;329(7471).
11. Moher D, Naudet F, Cristea IA, Miedema F, Ioannidis JPA, Goodman SN. Assessing scientists for hiring, promotion, and tenure. *PloS Biol*. 2018;16(3):e2004089.
12. Anderson MS, Ronning EA, De Vries R, Martinson BC. The perverse effects of competition on scientists' work and relationships. *Sci Eng Ethics*. 2007;13(4):437-61.
13. De Vries R, Anderson MS, Martinson BC. Normal Misbehavior: Scientists Talk About the Ethics of Research. *J Empir Res Hum Res Ethics*. 2006;1(1):43-50.
14. American Society for Cell Biology. San Francisco Declaration on Research Assessment. 2013. Available from: <https://sfdora.org/read/>.
15. Wilsdon J, Liz Allen, Belfiore E, Campbell P, Curry S, Hill S, et al. *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. HEFCE; 2015. Available from: <https://re.ukri.org/documents/hefce-documents/metric-tide-2015-pdf/>.
16. Hicks D, Wouters P, Waltman L, Rijcke Sd, Rafols I. The Leiden Manifesto for research metrics. *Nature News*. 2015;520:429-31.
17. Schekman R, Patterson M. Reforming research assessment. *Elife*. 2013;2:e00855.

18. Dijkstelbloem H, Huisman F, Miedema F, Mijnhardt W. Why science does not work as it should and what to do about it. Science in transition; 2013. Available from: <http://www.scienceintransition.nl/app/uploads/2013/10/Science-in-Transition-Position-Paper-final.pdf>.
19. Global Young Academy. Publishing models, assessments, and open science. Halle, Germany; 2018. Available from: <https://globalyoungacademy.net/wp-content/uploads/2018/10/APOS-Report-29.10.2018.pdf>.
20. Saenen B, Borell-Damián L. EUA Briefing - Reflections on University Research Assessment: Key concepts, issues and actors. European Universities Association; 2019. Available from: <https://eua.eu/resources/publications/825:reflections-on-university-research-assessment-key-concepts,-issues-and-actors.html>.
21. Science Europe. Position Statement and Recommendations on Research Assessment Processes. 2020 July. Available from: <https://scienceeurope.org/our-resources/position-statement-research-assessment-processes/>.
22. Moher D, Bouter L, Kleinert S, Glasziou P, Sham MH, Barbour V, et al. The Hong Kong Principles for assessing researchers: Fostering research integrity. PloS Biol. 2020;18(7):e3000737.
23. Curry S, Rijke Sd, Hatch A, Pillay DG, Weijden Ivd, Wilsdon J. The changing role of funders in responsible research assessment: progress, obstacles and the way ahead. Research on Research Institute; 2020 November. Available from: [https://rori.figshare.com/articles/report/The\\_changing\\_role\\_of\\_funders\\_in\\_responsible\\_research\\_assessment\\_progress\\_obstacles\\_and\\_the\\_way\\_ahead/13227914](https://rori.figshare.com/articles/report/The_changing_role_of_funders_in_responsible_research_assessment_progress_obstacles_and_the_way_ahead/13227914).
24. Latin American Forum for Research Assessment (FOLEC). Towards a Transformation of Scientific Research Assessment in Latin America and the Caribbean: Proposal for a Declaration of Principles. Latin American Council of Social Sciences (CLACSO); 2020. Available from: <https://www.clacso.org/en/una-nueva-evaluacion-academica-para-una-ciencia-con-relevancia-social-2/>.
25. Latin American Forum for Research Assessment (FOLEC). Towards a Transformation of Scientific Research Assessment in Latin America and the Caribbean: Evaluating Scientific Research Assessment. Latin American Council of Social Sciences (CLACSO); 2020. Available from: <https://www.clacso.org/en/una-nueva-evaluacion-academica-para-una-ciencia-con-relevancia-social/>.
26. Latin American Forum for Research Assessment (FOLEC). Towards a Transformation of Scientific Research Assessment in Latin America and the Caribbean: Diagnosis and Proposals for a Regional Initiative. Latin American Council of Social Sciences (CLACSO); 2020. Available from: <https://www.clacso.org/en/diagnostico-y-propuestas-para-una-iniciativa-regional/>.
27. European Open Science Cloud. Draft vision for FAIRer assessments. 2021. Available from: [https://avointiede.fi/sites/default/files/2021-02/eosc\\_cocreation\\_vision\\_for\\_fairer\\_assessments.pdf](https://avointiede.fi/sites/default/files/2021-02/eosc_cocreation_vision_for_fairer_assessments.pdf).
28. Ziker J. The Blue Review. 2014 31 March. Available from: <https://www.boisestate.edu/blureview/faculty-time-allocation/>.

29. Schimanski L, Alperin J. The evaluation of scholarship in academic promotion and tenure processes: Past, present, and future [version 1; peer review: 2 approved]. *F1000 Res.* 2018;7(1605).
30. Farrar J. Wellcome – News and views. 2019. Available from: <https://wellcome.ac.uk/news/why-we-need-reimagine-how-we-do-research>.
31. 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.* 2021;6(1):1.
32. Aubert Bonn N, Pinxten W. Advancing science or advancing careers? Researchers' opinions on success indicators. *PLoS ONE.* 2021;16(2):e0243664.
33. Kidwell MC, Lazarević LB, Baranski E, Hardwicke TE, Piechowski S, Falkenberg L-S, et al. Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PloS Biol.* 2016;14(5):e1002456.
34. EQUATOR network. Enhancing the QUALity and Transparency Of health Research [Available from: <https://www.equator-network.org>].
35. publons. [Available from: <https://publons.com/>].
36. ORCID. Connecting research and researchers [Available from: <https://orcid.org>].
37. CASRAI. CRediT – Contributor Roles Taxonomy [Available from: <https://casrai.org/credit/>].
38. Allen L, O'Connell A, Kiermer V. How can we ensure visibility and diversity in research contributions? How the Contributor Role Taxonomy (CRediT) is helping the shift from authorship to contributorship. *Learn Publ.* 2019;32(1):71-4.
39. European Commission. Academic Careers Understood through Measurement and Norms 2019 [updated 1 August. Available from: <https://cordis.europa.eu/project/id/266632>].
40. European Commission. Evaluation of Research Careers fully acknowledging Open Science Practices. Brussels, Belgium; 2017.
41. Hedding DW. Payouts push professors towards predatory journals. *Nature.* 2019;565(267).
42. Muthama E, McKenna S. The Unintended Consequences of Using Direct Incentives to Drive the Complex Task of Research Dissemination. *Education as Change.* 2020;24:23 pages.
43. Moed HF. UK Research Assessment Exercises: Informed judgments on research quality or quantity? *Scientometrics.* 2008;74(1):153-61.
44. Embassy of Good Science. Salami publication 2021 [Available from: [https://embassy.science/wiki/Theme:95c69cce-596a-42b5-9d86-e0aabaf00a85#Salami\\_publication](https://embassy.science/wiki/Theme:95c69cce-596a-42b5-9d86-e0aabaf00a85#Salami_publication)].
45. Hedding DW. Payouts push professors towards predatory journals. *Nature.* 2019;565:267.

46. Vogel L. Researchers may be part of the problem in predatory publishing. *Canadian Medical Association Journal*. 2017;189(42):E1324.
47. Eyre-Walker A, Stoletzki N. The Assessment of Science: The Relative Merits of Post-Publication Review, the Impact Factor, and the Number of Citations. *PloS Biol*. 2013;11(10):e1001675.
48. Cancer Research UK. 2018. Available from: <https://www.cancerresearchuk.org/funding-for-researchers/research-features/2018-02-20-improving-research-evaluation-dora>.
49. McKiernan EC, Schimanski LA, Muñoz Nieves C, Matthias L, Niles MT, Alperin JP. Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations. *Elife*. 2019;8:e47338.
50. Wikipedia. Impact factor [updated 7 July 2021. Available from: [https://en.wikipedia.org/wiki/Impact\\_factor](https://en.wikipedia.org/wiki/Impact_factor).
51. Lariviere V, Sugimoto CR. *Mesurer la Science*. Montréal, Canada: Les Presses de l'Université de Montréal; 2018. 169 p.
52. Garfield E. Der Impact Faktor und seine richtige Anwendung. *Der Anaesthesist*. 1998;47(6):439-41.
53. Hatch A, Schmidt R. Rethinking Research Assessment: Unintended Cognitive and System Biases. DORA; 2020. Available from: [https://sfdora.org/wp-content/uploads/2020/11/DORA\\_UnintendedCognitiveSystemBiases.pdf](https://sfdora.org/wp-content/uploads/2020/11/DORA_UnintendedCognitiveSystemBiases.pdf).
54. de Rijcke S, Wouters PF, Rushforth AD, Franssen TP, Hammarfelt B. Evaluation practices and effects of indicator use—a literature review. *Research Evaluation*. 2015;25(2):161-9.
55. Wouters P. The Citation: From Culture to Infrastructure. In: Cronin B, Sugimoto CR, editors. *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*. Cambridge (MA): The MIT Press; 2014. p. 47-66.
56. Ioannidis JPA, Thombs BD. A user's guide to inflated and manipulated impact factors. *European Journal of Clinical Investigation*. 2019;49(9):e13151.
57. Larivière V, Kiermer V, MacCallum CJ, McNutt M, Patterson M, Pulverer B, et al. A simple proposal for the publication of journal citation distributions. *bioRxiv*. 2016.
58. Brito R, Rodríguez-Navarro A. Evaluating research and researchers by the journal impact factor: Is it better than coin flipping? *J Inf*. 2019;13(1):314-24.
59. Schmidt R. Behavioral Scientist. 2020 24 August. Available from: <https://behavioralscientist.org/the-benefits-of-statistical-noise/>.
60. Baas J, Fennell C. When Peer Reviewers Go Rogue - Estimated Prevalence of Citation Manipulation by Reviewers Based on the Citation Patterns of 69,000 Reviewers ISSI 2019; 2-5 September 2019. Rome, Italy 2019.
61. Fong EA, Wilhite AW. Authorship and citation manipulation in academic research. *PLoS ONE*. 2017;12(12):e0187394.

62. Urlings MJE, Duyx B, Swaen GMH, Bouter LM, Zeegers MP. Citation bias and other determinants of citation in biomedical research: findings from six citation networks. *J Clin Epidemiol*. 2021;132:71-8.
63. Drake DC, Maritz B, Jacobs SM, Crous CJ, Engelbrecht A, Etale A, et al. The propagation and dispersal of misinformation in ecology: Is there a relationship between citation accuracy and journal impact factor? *Hydrobiologia*. 2013;702(1):1-4.
64. Hirsch JE. Superconductivity, What the h? The emperor has no clothes. *Physics and Society*. 2020;49(1):4-9.
65. Waltman L, van Eck NJ. The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology*. 2012;63(2):406-15.
66. Sugimoto CR, Larivière V. *Measuring research What everyone needs to know*. New York (NY): Oxford University Press; 2018. 149 p.
67. Garfield E. Is citation analysis a legitimate evaluation tool? *Scientometrics*. 1979;1(4):359-75.
68. Lindner MD, Torralba KD, Khan NA. Scientific productivity: An exploratory study of metrics and incentives. *PLoS ONE*. 2018;13(4):e0195321.
69. DORA. *Reimagining academic assessment: stories of innovation and change 2021* [Available from: <https://sfdora.org/dora-case-studies/>].
70. Open Science Policy Platform. OSPP-REC. Brussels, Belgium: European Commission; 2017. Report No.: ISBN 978-92-79-88333-0.
71. Winker K. Eyeballs on science: Impact is not just citations, but how big is readership? *bioRxiv*. 2017.
72. Alperin JP, Muñoz Nieves C, Schimanski LA, Fischman GE, Niles MT, McKiernan EC. How significant are the public dimensions of faculty work in review, promotion and tenure documents? *Elife*. 2019;8:e42254.
73. Lebel J, McLean R. A better measure of research from the global south. *Nature*. 2018;559:23-6.
74. Gingras Y, Mosbah-Natanson S. Les sciences sociales françaises entre ancrage local et visibilité internationale. *European Journal of Sociology*. 2010;51(2):305-21.
75. El-Sadr WM, Philip NM, Justman J. Letting HIV Transform Academia — Embracing Implementation Science. *New Engl J Med*. 2014;370(18):1679-81.
76. Leyser O. The excellence question. *Science*. 2020;370(6519):886.
77. Munafo MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie Du Sert N, et al. A manifesto for reproducible science. *Nature Human Behaviour*. 2017;1(1).
78. Bouter LM, Riet Gt. Replication Research Series-Paper 2 : Empirical research must be replicated before its findings can be trusted. *J Clin Epidemiol*. 2021;129:188-90.
79. Ioannidis JPA. Why replication has more scientific value than original discovery. *Behav Brain Sci*. 2018;41:e137.

80. Ofir Z, Schwandt T, Duggan C, McLean R. Research Quality Plus (RQ+) : a holistic approach to evaluating research. Ottawa, Canada: IDRC - CRDI; 2016. Contract No.: 10625/56528. Available from: <http://hdl.handle.net/10625/56528>.
81. Kirschner M. A Perverted View of “Impact”. *Science*. 2013;340(6138):1265.
82. Brooks R. Centuries wasted applying for grants? The Conversation. 2013 27 March. Available from: <https://theconversation.com/centuries-wasted-applying-for-grants-13111>.
83. Royal Society. Résumé for Researchers [Available from: <https://royalsociety.org/topics-policy/projects/research-culture/tools-for-support/resume-for-researchers/>].
84. Hatch A, Curry S. Changing how we evaluate research is difficult, but not impossible. *Elife*. 2020;9:e58654.
85. Larivière V, Ni C, Gingras Y, Cronin B, Sugimoto CR. Bibliometrics: global gender disparities in science. *Nature*. 2013;504(7479):211-3.
86. Beaudry C, Lariviere V. Which gender gap? Factors affecting researchers' scientific impact in science and medicine. *Research Policy*. 2016;45(9):1790-817.
87. Minello A. The pandemic and the female academic. *Nature*. 2020.
88. Viglione G. Are women publishing less during the pandemic? Here's what the data say. *Nature*. 2020;581:365-6.
89. Macaluso B, Lariviere V, Sugimoto T, Sugimoto CR. Is Science Built on the Shoulders of Women? A Study of Gender Differences in Contributorship. *Acad Med*. 2016;91(8):1136-42.
90. Astegiano J, Sebastián-González E, Castanho CdT. Unravelling the gender productivity gap in science: a meta-analytical review. *R Soc open sci*. 6(6):181566.
91. Check Hayden E. Racial bias continues to haunt NIH grants. *Nature*. 2015;527(7578):286-7.
92. Rochmyaningsih D. Showcase scientists from the global south. *Nature*. 2018;553:251.
93. Powell K. These labs are remarkably diverse — here's why they're winning at science. 2018;558,: 19-22.
94. Brock J. "Textbook case" of disability discrimination in grant applications. *Nature Index*. 2021 19 January. Available from: <https://www.natureindex.com/news-blog/textbook-case-of-disability-discrimination-in-research-grant-applications>.
95. Bothwell E. Award Nobels to teams, not individual ‘heroes’, say scientists. *Times Higher Education*. 2019 14 October. Available from: <https://www.timeshighereducation.com/news/award-nobels-teams-not-individual-heroes-say-scientists>.
96. Payne D. Calls for culture change as “them versus us” mindset drives rift between academic and non-academic staff. *Nature Index*. 2021 30 March. Available from: <https://www.natureindex.com/news-blog/calls-culture-change-them-versus-us-drives-rift-between-academics-administrators-research-science>.



97. Aubert Bonn N, Pinxten W. A Decade of Empirical Research on Research Integrity: What Have We (Not) Looked At? *J Empir Res Hum Res Ethics*. 2019;14(4):338-52.
98. Athena Swan Charter [Available from: <https://www.advance-he.ac.uk/equality-charters/athena-swan-charter>].
99. Dance A. How a Dutch university aims to boost gender parity. *Nature Career News*. 2019.
100. College oordeelt over voorkeursbeleid TU Eindhoven. College voor de Rechten van de Mens. 2020 3 July. Available from: <https://mensenrechten.nl/nl/nieuws/college-oordeelt-over-voorkeursbeleid-tu-eindhoven>.
101. IUPUI approves new path to promotion and tenure for enhancing equity, inclusion and diversity. *News at IUPUI*. 2021 10 May. Available from: <https://news.iu.edu/stories/2021/05/iupui/releases/10-promotion-tenure-pathway-enhancing-diversity-equity-inclusion.html>.
102. Labib K, Evans N. Gender, diversity, and the responsible assessment of researchers. *PloS Biol*. 2021;19(4):e3001036.
103. McNutt MK, Bradford M, Drazen JM, Hanson B, Howard B, Jamieson KH, et al. Transparency in authors' contributions and responsibilities to promote integrity in scientific publication. *Proceedings of the National Academy of Sciences*. 2018;115(11):2557.
104. Smith R. Authorship is dying: long live contributorship. *BMJ*. 1997;315:696.
105. Bone F, Hopkins MM, Ràfols I, Molas-Gallart J, Tang P, Davey G, et al. DARE to be different? A novel approach for analysing diversity in collaborative research projects. *Research Evaluation*. 2020;29(3):300-15.
106. Aubert Bonn N, Pinxten W. Rethinking success, integrity, and culture in research (part 2) — a multi-actor qualitative study on problems of science. *Research Integrity and Peer Review*. 2021;6(1):3.
107. Parr C. Open University maps new routes to career progression. *Times Higher Education*. 2015 2 April. Available from: <https://www.timeshighereducation.com/news/open-university-maps-new-routes-to-career-progression/2019410.article>.
108. Ghent University Department of Personnel & Organization. Vision statement and principles: New career path and evaluation policy for professorial staff. 2018 17 May. Available from: <https://www.ugent.be/en/work/mobility-career/career-aspects/professorial-staff/visionstatement.pdf>.
109. VSNU, NFU, KNAW, NWO, ZonMw. Room for everyone's talent. The Hague, The Netherlands; 2019 November. Available from: <https://vsnu.nl/recognitionandrewards/wp-content/uploads/2019/11/Position-paper-Room-for-everyone's-talent.pdf>.
110. Hoger Onderwijs Persbureau. Spinoza Prize to become a team effort. *Cursor*. 2019 7 October. Available from: <https://www.cursor.tue.nl/en/news/2019/oktober/week-2/spinoza-prize-to-become-a-team-effort/>.
111. Schneider SL, Ness KK, Shaver K, Brutkiewicz R. Federal Demonstration Partnership 2012 Faculty Workload Survey - Research Report. 2014. Available from:



- [https://osp.od.nih.gov/wp-content/uploads/SMRB\\_May\\_2014\\_2012\\_Faculty\\_Workload\\_Survey\\_Research\\_Report.pdf](https://osp.od.nih.gov/wp-content/uploads/SMRB_May_2014_2012_Faculty_Workload_Survey_Research_Report.pdf).
112. Aubert Bonn N, Pinxten W. Advancing science or advancing careers? Researchers' opinions on success indicators. bioRxiv. 2020:2020.06.22.165654.
  113. Herbert DL, Barnett AG, Clarke P, Graves N. On the time spent preparing grant proposals: an observational study of Australian researchers. BMJ Open. 2013;3(5):e002800.
  114. Rockey S. National Institutes of Health Office of Extramural Research. 2012 9 August. Available from: <https://nexus.od.nih.gov/all/2012/08/09/more-applications-many-more-applicants/>.
  115. Aubert Bonn N, Pinxten W. Rethinking success, integrity, and culture in research (Part 2) - A multi-actor qualitative study on problems of science. bioRxiv. 2020. doi:10.1101/2020.02.12.945899
  116. Gross K, Bergstrom CT. Contest models highlight inherent inefficiencies of scientific funding competitions. PloS Biol. 2019;17(1):e3000065.
  117. Smaldino PE, Turner MA, Contreras Kallens PA. Open science and modified funding lotteries can impede the natural selection of bad science. OSF Preprints. 2019. doi:10.31219/osf.io/zvkwq
  118. ISE task force on researchers' careers. Position on precarity of academic careers. Initiative for Science in Europe; 2020. Available from: <https://initiative-se.eu/wp-content/uploads/2021/02/Research-Precarity-ISE-position.pdf>.
  119. Kwon D. Prestigious European grants might be biased, study suggests. Nature News. 2021.
  120. Clauset A, Arbesman S, Larremore DB. Systematic inequality and hierarchy in faculty hiring networks. Science Advances. 2015;1.
  121. Hatch A. To fix research assessment, swap slogans for definitions. Nature. 2019;576(9):9.
  122. Vitae. 360 degree feedback from your research team [Available from: <https://www.vitae.ac.uk/doing-research/leadership-development-for-principal-investigators-pis/developing-yourself-as-a-pi/360-degree-feedback-from-your-research-team>].
  123. Hatch A, Curry S. Inside eLife [Internet]2019. Available from: <https://elifesciences.org/inside-elife/1fd1018c/research-assessment-reducing-bias-in-the-evaluation-of-researchers>.
  124. Health Research Board. HRB Gender Policy. 2019. Available from: [https://www.hrb.ie/fileadmin/user\\_upload/HRB\\_Gender\\_Policy\\_Nov\\_2019.pdf](https://www.hrb.ie/fileadmin/user_upload/HRB_Gender_Policy_Nov_2019.pdf).
  125. Nuffield Council of Bioethics. The culture of scientific research in the UK. 2014. Available from: <https://nuffieldbioethics.org/publications/the-culture-of-scientific-research>.
  126. The Wellcome Trust and Shift Learning. What Researchers Think About the Culture They Work In. 2020 15 January. Report No.: MC-7198/01-2020/BG. Available from: <https://wellcome.ac.uk/reports/what-researchers-think-about-research-culture>.

127. Janet Metcalfe, Katie Wheat, Munafo M, Parry J. Research integrity: a landscape study. 2020. Available from: <https://www.vitae.ac.uk/vitae-publications/reports/research-integrity-a-landscape-study>.
128. Larson RC, Ghaffarzadegan N, Xue Y. Too Many PhD Graduates or Too Few Academic Job Openings: The Basic Reproductive Number R0 in Academia. *Systems Research and Behavioral Science*. 2014;31(6):745-50.
129. Alberts B, Kirschner MW, Tilghman S, Varmus H. Rescuing US biomedical research from its systemic flaws. *Proceedings of the National Academy of Sciences*. 2014;111(16):5773.
130. van der Weijden I, Teelken C, de Boer M, Drost M. Career satisfaction of postdoctoral researchers in relation to their expectations for the future. *High Educ*. 2016;72(1):25-40.
131. Anonymous. The disposable academic. *The Economist*. 2010 December 18. Available from: <https://www.economist.com/christmas-specials/2010/12/16/the-disposable-academic>.
132. Many junior scientists need to take a hard look at their job prospects. *Nature*. 2017;550:429.
133. Debacker N, Vandeveld K. From PhD to professor in Flanders. *ECOOM Brief* (no 11). 2016. Available from: <https://biblio.ugent.be/publication/8043010>.
134. Martinson BC. The academic birth rate. Production and reproduction of the research work force, and its effect on innovation and research misconduct. *EMBO Rep*. 2011;12(8):758-61.
135. Padilla MA, Thompson JN. Burning Out Faculty at Doctoral Research Universities. *Stress and health : journal of the International Society for the Investigation of Stress*. 2016;32(5):551-8.
136. The mental health of PhD researchers demands urgent attention. *Nature*. 2019;575:257-8.
137. Levecque K, Anseel F, De Beuckelaer A, Van der Heyden J, Gisle L. Work organization and mental health problems in PhD students. *Research Policy*. 2017;46(4):868-79.
138. Heffernan TA, Heffernan A. The academic exodus: the role of institutional support in academics leaving universities and the academy. *Professional Development in Education*. 2019;45(1):102-13.
139. Van de Velde J, Levecque K, Mortier A, De Beuckelaer A. Waarom doctorandi in Vlaanderen denken aan stoppen met doctoreren [Why PhD students in Flanders think about stopping their PhDs]. *ECOOM Brief* (no 20). 2019 September. Available from: <http://hdl.handle.net/1854/LU-8630419>.
140. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. 2016;3(1):160018.
141. Cogitamus Laboratory. Camille Noûs 2020 [Available from: <https://www.cogitamus.fr/camilleen.html>].
142. University and College Union. 2020 3 February. Available from: <https://www.ucu.org.uk/article/10621/UCU-announces-14-strike-days-at-74-UK-universities-in-February-and-March>.

143. Regeringskansliet. Forskning, frihet, framtid – kunskap och innovation för Sverige. Sweden; 2020 17 December. Contract No.: 2020/21:60.
144. Regeringskansliet. Statens offentliga utredningar från Utbildningsdepartementet: En långsiktig, samordnad och dialogbaserad styrning av högskolan. Sweden; 2019 1 februari. Contract No.: SOU 2019:6. Available from: <https://www.regeringen.se/rattsliga-dokument/statens-offentliga-utredningar/2019/02/sou-20196/>.
145. Hwang S. Forskningskvalitet, effektivitet och extern finansiering. Sweden: Halmstad University Press; 2018. Report No.: ISBN 978-91-88749-06-2. Available from: <http://hh.diva-portal.org/smash/record.jsf?pid=diva2%3A1253091&dswid=-5182>.
146. Mejlgaard N, Bouter LM, Gaskell G, Kavouras P, Allum N, Bendtsen A-K, et al. Research integrity: nine ways to move from talk to walk. 2020;586:358-60.
147. Jonkers K, Zacharewicz T. Research Performance Based Funding Systems: a Comparative Assessment. European Commission; 2016. Report No.: ISBN 978-92-79-68715-0 - ISSN 1831-9424. Available from: <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC101043/kj1a27837enn.pdf>.
148. Debackere K, Glänzel W. Using a bibliometric approach to support research policy making: The case of the Flemish BOF-key. *Scientometrics*. 2004;59(2):253-76.
149. Engels TCE, Guns R. The Flemish Performance-based Research Funding System: A Unique Variant of the Norwegian Model. *Journal of Data and Information Science*. 2018;3(4):45-60.
150. Gadd E. University rankings need a rethink. *Nature*. 2020;587(523).
151. Hazelkorn E. The Impact of League Tables and Ranking Systems on Higher Education Decision Making. *Higher Education Management and Policy*. 2007;19(2).
152. Lariviere V, Sugimoto CR. *Vue d'ensemble. Mesurer la Science*. Montréal, Canada: Les Presses de l'Université de Montréal; 2018. p. 145-62.
153. Universities Norway. NOR-CAM - A toolbox for recognition and rewards in academic careers. Oslo; 2021 May. Available from: [https://www.uhr.no/en/\\_f/p3/i86e9ec84-3b3d-48ce-8167-bbae0f507ce8/nor-cam-a-tool-box-for-assessment-and-rewards.pdf](https://www.uhr.no/en/_f/p3/i86e9ec84-3b3d-48ce-8167-bbae0f507ce8/nor-cam-a-tool-box-for-assessment-and-rewards.pdf).
154. Working group for responsible evaluation of a researcher. Good Practice in Researcher Evaluation. Recommendation for the Responsible Evaluation of a Researcher in Finland. Helsinki: The Committee for Public Information (TJNK) and Federation of Finnish Learned Societies (TSV); 2020. Report No.: 978-952-5995-28-2. Available from: <https://doi.org/10.23847/isbn.9789525995282>.