Kramer, Bianca; Bosman, Jeroen

Book Part

Chapter 4 Recognition and Rewards in Academia – Recent Trends in Assessment

Reference: In: Talent Management in Higher Education (2024). Emerald Publishing Limited, S. 55 - 75.
https://doi.org/10.1108/978-1-80262-685-820241004.
doi:10.1108/978-1-80262-685-820241004.

Chapter 4

Recognition and Rewards in Academia – Recent Trends in Assessment

Bianca Kramer[a] *and Jeroen Bosman*[b]

[a]*Sesame Open Science, The Netherlands*
[b]*Utrecht University Library, The Netherlands*

Abstract

In academia, assessment is often narrow in its focus on research productivity, its application of a limited number of standardised metrics and its summative approach aimed at selection. This approach, corresponding to an exclusive, subject-oriented concept of talent management, can be thought of as at odds with a broader view of the role of academic institutions as accelerating and improving science and scholarship and its societal impact. In recent years, open science practices as well as research integrity issues have increased awareness of the need for a more inclusive approach to assessment and talent management in academia, broadening assessment to reward the full spectrum of academic activities and, within that spectrum, deepening assessment by critically reflecting on the processes and indicators involved (both qualitative and quantitative). In terms of talent management, this would mean a move from research-focused assessment to assessment including all academic activities (including education, professional performance and leadership), a shift from focus on the individual to a focus on collaboration in teams (recognising contributions of both academic and support staff), increased attention for formative assessment and greater agency for those being evaluated, as well as around the data, tools and platforms used in assessment. Together, this represents a more inclusive, subject-oriented approach to talent management. Implementation of such changes requires involvement from university management,

Talent Management in Higher Education, 55–75

doi:10.1108/978-1-80262-685-820241004

human resource management and academic and support staff at all career levels, and universities would benefit from participation in mutual learning initiatives currently taking shape in various regions of the world.

Keywords: Academic culture; agency; culture change; evaluation process; impact; metrics; qualitative indicators; quantitative indicators; research assessment

The Role of Assessment in Shaping Academia

For institutions that include acceleration and improvement of science and scholarship as well as their societal impact in their mission, it is important to use assessment of academic activities and their outcomes that align with those missions.

Assessment of academic activities, both within and outside universities, determines how research budgets are allocated and who is hired, promoted and given tenure. It also plays a role in how the university is viewed as a partner for international collaborations, and whether potential new students and employees view the university as a desirable place to study and work. Therefore, universities are strategic in how they want to be perceived compared to other institutions, both nationally and internationally.

This has a direct link with talent management – the way a university thinks about talent and how to best attract and sustain it. In literature on talent management, a distinction is made between talent conceived of as subject (with a focus on people as 'talents') and talent viewed as object (where 'talents' are characteristics of people, such as abilities, knowledge and/or competencies). In addition, a distinction is made between exclusive and inclusive talent management, with exclusive talent management focusing on selection of people or characteristics at the exclusion of others, and inclusive talent management as more broadly considering the need for multiple qualities to support the organisation's overall objective (Thunnissen et al., 2013). How a university approaches talent management has a direct relationship with how assessment is taking place.

This raises questions as to how to shape assessment of academic activities to promote the academic culture that research organisations aspire to, internally as well as for the higher education system as a whole. Choices in assessment of academic activities (whom to assess, what to assess and how to assess) have the potential to shape both the institution and the wider system of higher education and research, and care should be taken to align assessment practices with the core values of the institution and the system as a whole. In essence, then, the question becomes: what kind of institution do universities want to be? From that, choices in assessment practices follow.

Importantly, this way of thinking also provides a key to change when current assessment practices do not support these core values – when certain essential academic activities and roles are undervalued compared to others, when a focus

on competition fosters a culture of individualism rather than teamwork and collaboration and when success is defined by narrow measures of quality and impact rather than reflect the multiple qualities of academia (Advisory Council for Science, Technology and Innovation (AWTI), 2023) that together provide true relevance for science and society.

The previous chapters of this book have described how the academic landscape has changed and how open science is altering academic tasks, systems and structures. In this chapter, we will discuss what this means for assessment, by looking at current developments in the Netherlands and internationally. What choices can be made by a university in how hiring, promotion and tenure decisions are made, both for academic and non-academic staff? What are the issues with commonly used metrics for academic success, and what alternative approaches are being proposed?

This chapter will first discuss the concept of both broadening and deepening assessment: rewarding the full spectrum of academic activities and, within that spectrum, critically reflecting on the process of quality and impact assessment. This will be followed by a closer look at the role of qualitative and quantitative assessment and appropriate use of indicators in both. Next, the relation that open science and research integrity play in assessment will be discussed, as well as the importance of equity and open infrastructure, with a special look at university rankings. Finally, some examples will be given of how changes in assessment practices are implemented at different academic institutions in the Netherlands, and a number of actionable international developments will be highlighted that could provide a springboard for further action.

Broadening and Deepening Assessment

For assessment practices to optimally support the role of universities to accelerate and improve science and scholarship and its societal impact, two aspects are important. First, they should reward the full spectrum of academic activities and not focus primarily on research. Second, they should reward practices that improve the quality, relevance and impact of academic activities, using appropriate indicators and processes. These two aspects can be conceived of as 'broadening' and 'deepening' assessment.

Traditional assessment is often relatively narrow, with its limitation to research and within that to (journal) publications. It is also often relatively shallow, with the orientation at measurable output, the importance of quantity and the use of a small number of standardised metrics. Aubert Bonn and Bouter (2023) describe how metrics use in assessment developed from mere quantitative measurement to impact measurement through citation counts and journal impact factors (JIFs) but also how both provided incentives that could harm research. In addition, traditional assessment often uses a comparative–summative approach aimed at selection (Aguinis et al., 2020; Kallio et al., 2017). Finally, it often employs 'excellence' as its central tenet (Moore et al., 2017). The concept of 'excellence', while difficult to define, is used by many institutions and underpins approaches that are

highly selective and often based on proxies for quality, such as journal metrics or lists of approved journals. Obviously, the exact set of criteria used differs between institutions and between the various assessment contexts such as hiring, tenure and promotion, grant allocation, prizes, etc. (see, e.g., McKiernan et al., 2019; Moher et al., 2018). In terms of talent management, this approach to assessment corresponds to an exclusive, subject-oriented approach, Thunnissen et al. (2021) focused on individual performance appraisal (Boselie, 2014). In contrast, broadening and widening assessment can be seen as moving towards a more inclusive, object-oriented approach.

Broadening Assessment

The concept of broadening assessment of academic activities means acknowledging that for a university to meet the expectations set upon it, more is needed than high-quality and relevant research. Education, for one, forms a large part of what a university is, and to do it well, it should be recognised and rewarded as an academic activity at par with doing research, with enough time, resources and recognition allocated to it. Many academic activities also fall under 'professional performance', be it clinical work at university hospitals or university veterinary hospitals, serving on governing or advisory boards of professional societies or associations, editorial work for scholarly journals and books, to name but a few activities. Another important area is leadership: time invested in managing a research group, fostering an open and inclusive research culture and mentoring trainees are important activities, relevant in all stages of an academic career. Two other important aspects of leadership are taking the lead and responsibility in innovation and improvement of processes and services and personal leadership: self-reflection in order to perform well.

It should be apparent that these activities require dedicated time and skills and thus need to be recognised and rewarded as valuable activities on their own. When academics are primarily valued for their research activities and outcomes, but at the same time are expected to carry out these other tasks as well, this can result in overburdening people who are expected to do it all, or creating 'second-class citizens' within academia, e.g., when teaching is performed by people on temporary contracts who have less favourable career opportunities within academia.

A corollary of the above is that no single person can or should be expected to excel at all academic activities or, in other words, be the elusive 'sheep with five legs'. Rather, success in academia is a team effort, and recognising this in assessment opens the door for more diverse career paths that are considered equally valuable. It also makes it easier to value contributions by support staff as bona fide academic activities – including, but not limited to, activities of lab technicians, data stewards, research software engineers, librarians who often work closely together with research groups and contribute to research, teaching and professional performance. Recognising these contributions also fits with assessment at team level, where the focus is on the functioning of the team as a whole and the contributions of all team members.

Ultimately, research, education and professional performance, supported through leadership and team science, result in impact. This can be either scientific or scholarly impact (e.g., contributions to theory, methodology or results leading to practical applications) or societal impact (e.g., practical applications, contributions to societal discourse, public–private partnerships). Here, too, broadening the concept of impact in assessment is important to value diverse academic activities more equitably, rather than focus on a narrow sense of research impact as most valuable, or consider 'impact' as societal impact only, separate from impact from research or education.

In the Netherlands, examples of the idea of broadening assessment can be found in the joint position paper 'Room for everyone's talent' (VSNU, 2019) and the implementation of the ambitions expressed therein at each individual university. Utrecht University has operationalised the concept of TRIPLE (Fig. 4.1) with Research, Education and Professional Performance supported and enabled by Team Science and Leadership as the scaffolding for broadening assessment (Utrecht University, 2021). This has since been implemented in requirements for tenure and promotion, as well as in the template for performance review for both academic and non-academic staff. For the latter, a decision could be made to rename the three top leaves of the lotus flower model (research, education, professional performance) to reflect relevant other task domains, such as administration, information technology services or facility management.

For hiring and function profiles, the broadening of assessment can also translate into describing various types of academic functions, based on different profiles, with different sets of tasks and requiring different sets of skills and expertise. For instance, the University Medical Centre Utrecht (UMCU), an early mover in assessment reform, has introduced six academic career profiles: clinical researcher, academic educator, exploration researcher, implementation researcher, methodology and technology researcher and valorisation researcher (UMCU, 2022, 2023).

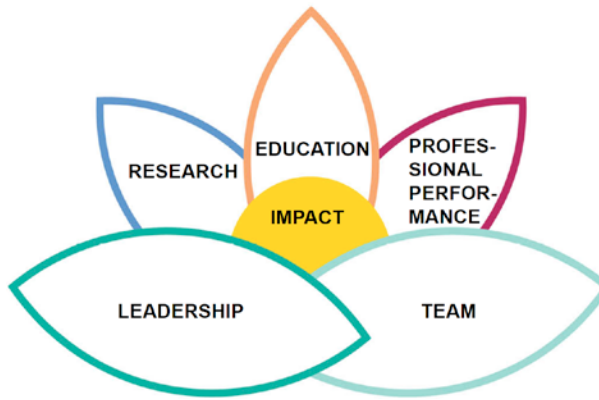

Fig. 4.1. TRIPLE Model for Recognition and Rewards, Utrecht University. *Source*: Provided courtesy of Utrecht University.

Deepening Assessment

As discussed above, broadening assessment from a narrow focus on research outcomes to a wider valuation of academic activities and contributions can help reduce pressure on single individuals to 'do everything' while being assessed primarily on research outcomes and can stimulate diversity in academic career paths by explicitly valuing all academic activities. In itself, though, this is no guarantee that for any given type of academic activity, quality and impact are appropriately assessed and, through that, encouraged.

An important question then is: what are appropriate indicators for the broadening assessment of academic activities? The risk in this context is the use of proxy indicators, particularly quantitative indicators, for quality and impact. Well-known examples are the use of the JIF in research assessment as a proxy for both quality and impact or article-level citations as a proxy for quality (McKiernan et al., 2019). There are a number of risks associated with the use of such proxy indicators:

- the risk that the proxy does not measure what it is intended to measure (methodological risk);
- the risk that the proxy is used primarily because of its availability, not because of its relevance or methodological quality (streetlight effect);
- the risk of the proxy being used in isolation, without taking into account other indicators; and
- the risk that the proxy indicator, rather than the underlying quality, becomes the target-guiding practices of both academics and organisations (Goodhart's law).

Taken together, the uncritical use of a limited set of proxy indicators can lead to perverse incentives (Hicks et al., 2015). This will be further elaborated on in Section 3 (Quantitative or Qualitative Assessment). In this section, we will address a number of approaches to move beyond such a limited approach. While we draw our examples primarily from research assessment as we are most involved and familiar with this domain, the same considerations and approaches are relevant for education and professional performance.

One approach is to critically reflect on the indicators used: are they appropriate indicators for the purpose for which they are used? Are there other indicators that can complement or even replace the indicators used, e.g., to look at a broader set of outputs, and include a broader set of indicators for relevance and impact? For research assessment, this approach has been advocated by the San Francisco Declaration of Research Assessment (DORA) which recommends

> For the purposes of research assessment, consider the value and impact of all research outputs (including datasets and software) in addition to research publications, and consider a broad range of impact measures including qualitative indicators of research impact, such as influence on policy and practice. (DORA, 2013)

Another aspect of 'deepening assessment' is to not look solely at outputs (be it of research, education or professional performance) but consider the activities and processes (including leadership and teamwork) leading to these outputs as subjects of assessment. One benefit of this is that assessment can be more formative: asking groups or individuals to reflect on their strategic goals and the activities undertaken to achieve these goals, as well as on the results thereof. In this way, assessment can bring about changes in process going forward, rather than be a reflection of 'success' or 'failure' after the fact. This is the approach taken by the Dutch Strategy Evaluation Protocol (VSNU, 2020) (formerly the 'Standard Evaluation Protocol', a telling rebranding it itself) in shaping the periodic formative evaluation of research groups. It allows for context-specific choices (accounting for disciplinary differences) as well as for choices to be made collectively at the level of the groups (research groups, departments, faculties) being evaluated.

In addition, focusing on process shifts the focus from producing outputs (which can itself be a perverse incentive) to safeguarding good processes. Fig. 4.2 provides an overview of possible processes to include in research evaluation, with accompanying aspects that could be considered. Similar thought exercises could be envisioned for, e.g., education and professional performance.

Finally, it is important to critically reflect on *who* sets the criteria for what is included in assessment and *how* assessment takes place. Especially when strategic goals are the starting point, the activities and outputs included in assessment, as well as any indicators used, should ideally be decided on in dialogue with who is assessed, rather than be decided for them. In addition, it should be carefully considered whether indicators used are appropriate for both the 'aggregation level' at which they are used and the goal of assessment.

Assessment can take place at various levels: the individual, a research group or department, a university as a whole or even a whole country. It has already

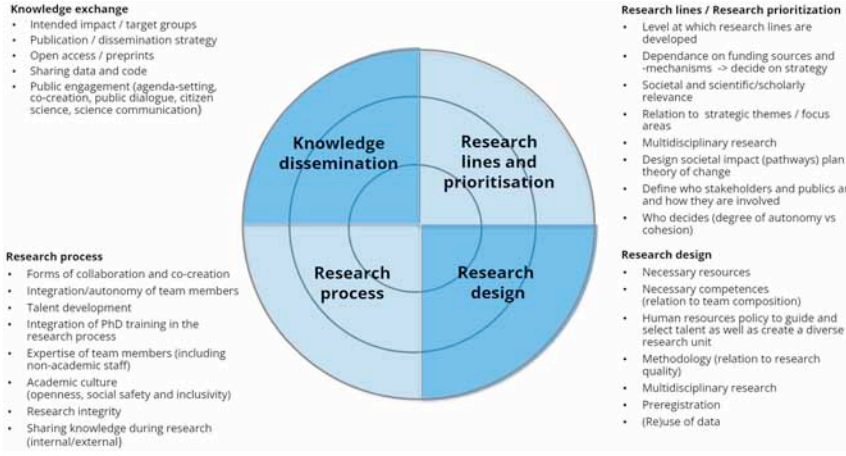

Fig. 4.2. Various Aspects of the Research Workflow That Could Be Considered in Assessment to Focus on Process Rather Than Outcomes.

been discussed how a focus away from individuals and towards teams can allow for more diverse career paths and recognition of a wider spectrum of competencies (also reflecting a more inclusive, object-oriented approach to talent management). Assessment at the level of institutions or countries usually has a different role, more focused on comparing performance or understanding the effects of local differences.

Depending on both the level at which assessment is taking place, and the goal of assessment, the use of certain indicators may not be appropriate. For example, institutional-level indicators do not reflect the qualities of individual people working or studying at that institution. Similarly, using indicators with the goal to increase understanding or even for promotional purposes carries lower risks for the entities being assessed than the use of metrics for incentivising or deciding on distribution of rewards (Gadd, 2019).

Finally, assessment could also take into account various forms of hybridity, especially when institutions apply their renewed assessment goals and process to all staff, including support staff. Hybridity might involve people having mixed functions (e.g., part-time in an academic role and part-time in a support role), people switching between roles during their career (e.g., an academic moving into research policy for a few years and then back into research and teaching) and people being part of mixed project teams consisting of academic as well as support staff. All three have repercussions for choosing assessment criteria and for the design of the process, in particular the question who is involved in assessing.

Qualitative or Quantitative Assessment

One question that has been getting a lot of attention in the discussion around research assessment is the role of metrics versus peer review, sometimes put as a dichotomy between quantitative and qualitative evaluation. Peer review, defined by the European University Association (EUA) as the process of experts making a qualitative judgement of research quality (Saenen & Borrell-Damián, 2019), refers to the process where one or more individuals perform in-depth assessment, often followed by a consultation between the peer reviewers, or a comparison and synthesis of their assessments. Peer review can take place at various levels, both for assessing individual outputs (like research articles undergoing peer review before being published in a journal or grant proposals being assessed for funding), assessing individuals (for hiring tenure and promotion) and assessing research groups (like in the Dutch Strategy Evaluation Protocol (VSNU, 2020) which involves site visits).

Peer review is sometimes considered the ‘gold standard’ – assuming assessment by a group of peers with knowledge of a specific discipline and context, can be expected to be more reliable than relying on metrics to make the ‘right’ decisions on, e.g., grant allocation, benchmarking research groups or hiring and promotion decisions. However, peer review has been shown to carry substantial variety in judgement between experts (peer reviewers) (Bertocchi et al., 2015; Cole et al., 1981; Traag & Waltman, 2019), raising questions on whether any decisions on, e.g., grant proposals objectively reflect the ‘right’ outcome, and even whether

such an objectively right outcome exists in the first place (Lee et al., 2013). Other arguments against peer review that can be made are its sensitivity to subjective decisions (Teplitskiy et al., 2018), including the phenomenon that that search and hire commissions often gravitate towards candidates who are similar to them, as they are looking for a good 'fit' – leading to a lack of diversity (van den Brink & Benschop, 2014), as well as the argument that more qualitative methods often associated with peer review take a lot of time and are therefore sometimes considered unsustainable (Bendiscioli, 2018; Singh Chawla, 2019).

In practice, assessment decisions made through peer review often already include the use of metrics or other indicators as part of the information gathered, and therefore, there is less of a dichotomy between qualitative and quantitative assessment, and more a question of what indicators are suitable for use in a given context. Also, both qualitative-based (review) and quantitative-based (metric) assessment reports can come with contextualisation and interpretation. As mentioned above, a few aspects to consider here are a) the validity of an indicator for the purpose it is used for; b) avoiding the 'streetlight' effect or choosing an indicator because it is available, rather than because it is the most appropriate; c) allowing a variety and diversity of indicators, rather than one or two default ones; and d) the risk that the indicator itself becomes a target (Goodhart's law).

(Not) Fit for Purpose

These aspects are all in play in cases where it is customary to use indicators which are not fit for purpose but which are used because they are readily available and commonly used by others and where there is resistance, distrust or just uncertainty towards using more diverse and less standardised indicators.

Two examples of this are the use of the JIF and the h-index in assessing research quality. The JIF is a metric at the level of an academic journal, giving (roughly speaking) the average number of citations in a given year to papers published in the journal in the two preceding years. There are a number of issues with the use of JIF to assess research quality, both methodologically and conceptually (for a summary, see Larivière & Sugimoto, 2019, and Fig. 4.3). Arguably, the most important one is that, being an average at the journal level, the JIF does not reflect or predict the number of citations to any given (published or future) paper, as illustrated by the observed skewness of citation distribution in many journals (Larivière et al., 2016). A second, more general argument is that citations in themselves do not necessarily reflect quality and only reflect a particular type of impact (i.e., used as reference by other academics). Despite this, the JIF is used so ubiquitously in evaluations that publishing in high-impact journals has become a target in itself that shapes research practice (a prime example of Goodhart's law).

Somewhat similarly, the h-index (most often a metric at the level of individual researchers but which could also be applied at the level of journals or institutions) reflects the number of x publications (e.g., of a specific author) that each have received x or more citations. As such, it is a metric that favours late career over early career researchers (as the h-index can only rise over the course of a career) and again only reflects citations as at best a narrow metric of quality and impact.

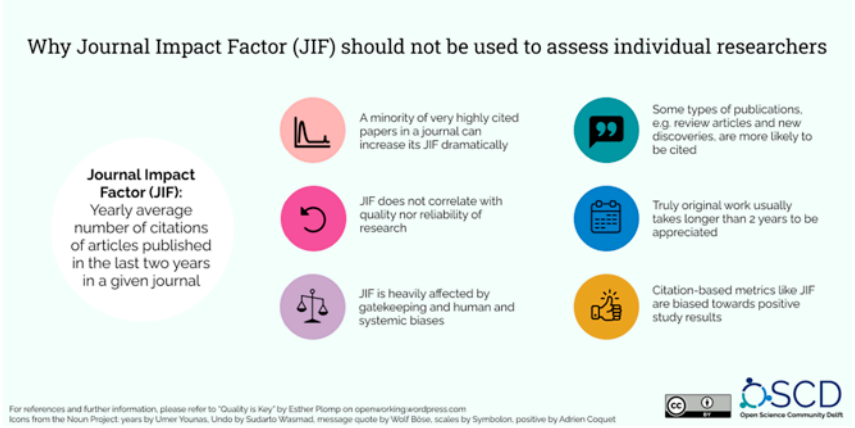

Fig. 4.3. Why JIF Should Not Be Used to Assess Individual Researchers (Plomp et al., 2021). Image license: CC-BY.

For an overview of the discussions around other problematic aspects of the use of the h-index, see Bornmann and Daniel (2009) and de Rijcke et al. (2021) and also Fig. 4.4.

A compelling visual example of the various types of usage and impact of a researcher's output and activities that are disregarded when a narrow focus on JIF, h-index and citations in general is applied is provided in the infographic 'I am not my h-index (or my JIFs)' (Fig. 4.5), where against a background of a simple plot of number of publications and number of citations, a number of publications are highlighted with the specific impact they have had.

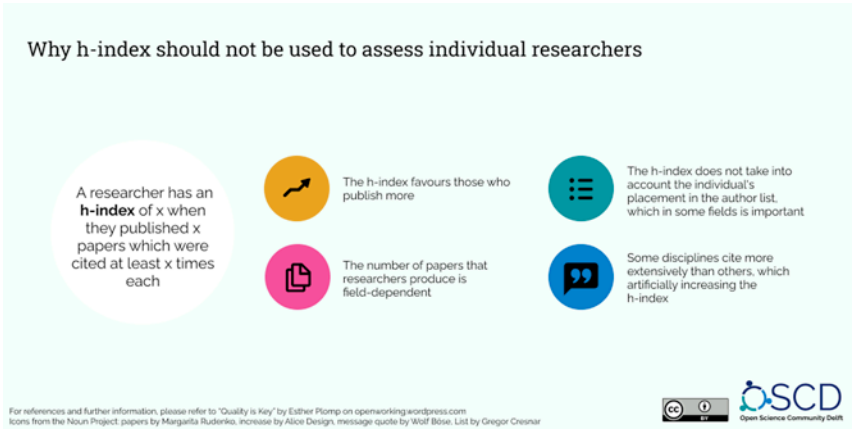

Fig. 4.4. Why *h*-Index Should Not Be Used to Assess Individual Researchers (Plomp et al., 2021). Image license: CC-BY.

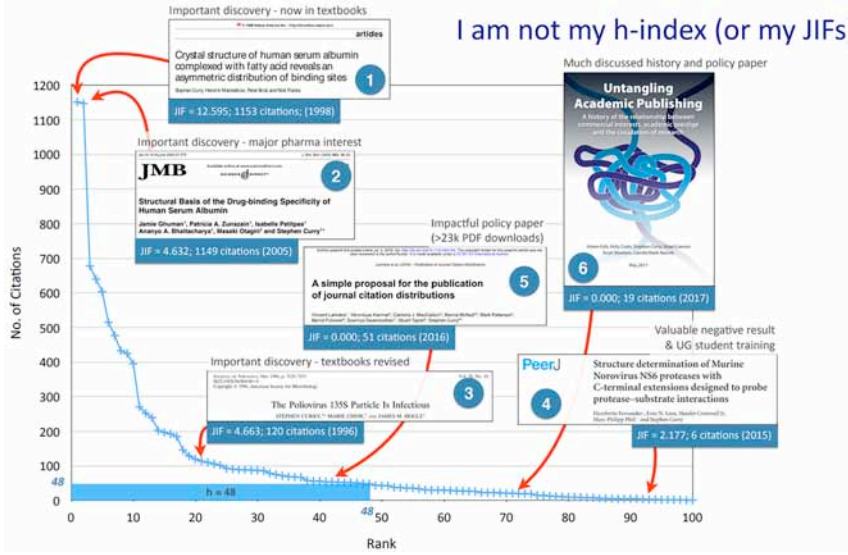

Fig. 4.5. I Am Not My *h*-Index (or My JIFs) (Curry, 2018). Image license: CC-BY.

Alternative Approaches

At the surface level, examples of the use of broader indicators for research assessment include looking at indicators for societal impact (e.g., use of research in policy documents and public debate), looking at citation diversity, rather than citation counts, as a measure for the (academic) audience reached (Huang et al., 2022). At a deeper level, leaving the decision on which outputs and activities to report on, and which indicators to provide to demonstrate quality and impact, up to the individual researcher is an approach taken by the Dutch Research Council (NWO) in a number of its funding schemes (Dutch Research Council (now), n.d.; Gossink-Melenhorst, 2019). It is echoed in the Dutch Strategy Evaluation Protocol (VSNU, 2020) for the periodic evaluation of research groups at Dutch universities, which is aimed at aligning research evaluation with the aims and goals of those being assessed, rather than with a standardised concept of what counts as good performance. Finally, the explicit guidance provided by NWO to not use aggregate indicators (like JIF) to provide evidence of quality and impact of individual research outputs is a prime example of deepening research assessment – addressing inappropriate use of specific indicators.

NWO also implemented the narrative, or evidence-based, CV – a combination of a narrative section to showcase the candidate's expertise and experience relevant to the project, and a 'key output' section to list a maximum of 10 outputs (not necessary publications) and indicators for their quality and impact. Other funders and institutions are also introducing the concept of narrative or evidence-based curriculum vitae as a way to enable qualitative and quantitative

assessment in a contextualised way (Woolston, 2022), with the choices of what to present driven by those assessed rather than by those doing the assessment. It is important to stress that these formats still include objective indicators that can be assessed for their relevance and value by those performing assessments. Nonetheless, responses to narrative CVs have been mixed (Bordignon et al., 2023), with often-heard criticism that it favours those with the ability to 'sell themselves' on paper and fear that it will make assessment more subjective. Kaltenbrunner et al. (2023) have proposed a research agenda to get a better sense of the extent to which narrative CVs can be effective as part of a coordinated broader strategy to foster inclusive practices in research evaluation and of the practical conditions that must be met to achieve this potential.

More in general, the right balance between relevance and contextualisation, on the one hand (with more options for those being assessed to select what to present to assessors), and comparability, on the other hand (with more standardised requirements for outputs and metrics to report and use), needs to be decided for any given assessment exercise. This requires critical reflection on the actual goal of the specific assessment, the capacity of reviewers/assessors to engage with a variety of reported outputs and indicators to compare candidates and any changes needed at the organisational level to enable broader assessments. Whatever the process, care should be taken not to default to the use of a narrow set of metrics only for convenience, especially where there are concerns about their validity for the type of assessment at hand.

Finally, it is good to realise that the need for standardisation also depends on the kinds of comparisons one wishes to support: are these only with one's own goals or previous performance? Or is comparison needed within an institution, or even nationally or globally? The wider the required scope of comparison, the stronger the need is for standardisation of criteria used. For cases where wide comparability is required, it remains the question to what extent it is possible to have assessments that combine standardised criteria with more variable types of evidence.

Open Science and Research Integrity

Two aspects of academic assessment that are associated with both broadening and deepening assessment are open science and research integrity. When open science is seen as open sharing of research processes and outputs (like data and software code) and open reporting (open access to publications and using open peer review), including it in assessment can be as straightforward as broadening assessment to include these outputs and activities, as well as recognising time and effort required to make research output openly available. This could also include recognising the roles of, e.g., data stewards and research software engineers in the research process, as part of a broader recognition of team science. When open science is considered more broadly as also including outreach activities, these activities could be similarly recognised in assessment. Recognising open science in this way helps create a research environment in which these activities are not seen as taking time away from but as integral part of doing research. Taking this

one step further, considering from the start what parties are relevant stakeholders both in setting a research agenda and in making use of the outcomes of research, can help set a publication and dissemination strategy that is optimised to reach these goals and which in turn can be the basis of evaluation and assessment. In such an integrated approach, openness can be considered a goal, rather than a characteristic of research outputs only.

Open science not only makes more parts of the research process accessible to a wider audience, it also makes the research process more transparent and, as such, contributes to research integrity. This includes making protocols, data and code openly available, including through preregistration. Stimulating openness to this end can help to prevent questionable research practices and promote responsible ones (Gopalakrishna, Ter Riet, et al., 2022; Gopalakrishna, Wicherts, 2022) and match expectations set on researchers by research integrity guidelines and mandates (see, for example, VSNU, 2018, Universities UK, 2019). To achieve this, it is important to ensure such openness, as well as the time and effort required to achieve it, is recognised accordingly in research evaluation (Bouter, 2023). With that also comes the recognition that the possibility for error detection that is facilitated through transparency is not just aimed at detecting potential scientific fraud but primarily at improving the scientific process, e.g., by increasing the chances of catching inadvertent errors in code and data and checking the robustness of results through reproducibility and replication studies. For this, an error-friendly environment is paramount, where errors are not seen as a stigma but as part of the process to improve science.

There are two additional important considerations regarding the role of open science and research integrity in research evaluation. First, there is no one-size-fits-all approach in what practices to encourage and/or require, including through research evaluation. Not all aspects of open science are relevant or viable for all disciplines or, within disciplines, for all research projects. A prime example is the restriction on data sharing imposed by privacy concerns (including compliance with the General Data Protection Regulation (GDPR). This does not mean that open science is not relevant in these cases, but that the aspects of open science that are relevant and feasible, and the ways these are implemented, often are dependent on the specific context of the research. Acknowledging this in research evaluation can encourage adoption of open science practices without it being perceived as a straight jacket.

Second, research evaluation also has the potential to shape research practices through what is seen as valued and needed for career advancement. In that sense, including open science in research evaluation can be seen as not just rewarding additional practices but shifting the focus from research outcomes (rewarding novelty and attention) to research process, rewarding robust methodology and transparency in reporting. This is another view on what high-quality research means.

Equity, Open Infrastructure and Rankings

So far, we have discussed both broadening and deepening research assessment – including a wider range of academic activities, outputs and indicators for quality

and impact and critically assessing the validity of the indicators used. One additional aspect to question is the tools and platforms providing information on both research outputs and indicators for quality and impact, as this determines both what is included and excluded and who is given deciding power over this. Many universities manage a current research information system (CRIS) to maintain a record of research outputs and/or make use of commercial, often off-the-shelf, tools for research metrics, like SciVal (provided by Elsevier) or Incites (provided by Clarivate) that are built on commercial bibliographic databases (Scopus and Web of Science, respectively). There are several limitations to the use of proprietary data for research assessment. First, where the coverage of such databases is selective, where selectively is decided by the (commercial) provider, this transfers agency away from the research organisation or individual researchers to the data provider. In practice, this can mean research outputs are limited to journal articles only, skewed to English language publications from Western countries and restricted to journals that meet the providers' definition of quality and impact. Especially when the desire is to broaden research assessment to more varied research outputs and indicators, this can be problematic. Compounding this issue, research information in commercial proprietary databases cannot be accessed or shared publicly, so can be neither verified or have a different lens applied to by others, including the ones being assessed. In other words, assessment and talent management are affected by the companies providing the assessment criteria and data, which raises issues with respect to distributive and procedural justice.

In recent years, the availability of openly available research information (open for both access and reuse) has grown, with initiatives such as Open Citations (https://opencitations.net), the OpenAIRE Research Graph (https://graph.openaire.eu/) and most recently OpenAlex (https://openalex.org/) providing alternatives to closed, proprietary bibliographic or citation databases. Use of these platforms, especially by public institutions, not only allows institutions more control over the selection of data used for assessment but also supports the idea of these type of data to be managed through public, rather than commercial infrastructures, in line with recommendations such as the UNESCO Recommendation on Open Science (UNESCO, 2021) and the COARA Agreement on Reforming Research Assessment (COARA, 2022). Where the use of open and/or non-proprietary sources would require additional time investment (e.g., to clean data and write code to analyse the data), this could be taken up in the form of collaborations of research performing and research funding organisations, including consortial funding of non-profit infrastructure that utilises open data.

Finally, a word on rankings. An argument that is often made against initiatives for broadening and broadening research assessment is that research organisations cannot decide to change their evaluation systems unilaterally, as in order for them to be attractive for potential employees and students, as well as give their current employees and students the best changes for future study and employment at other institutions, conforming to commonly held ideas on quality and how to assess it is important. Nowhere is this more important as in the ambivalence around university rankings. There are serious reservations with both the methodology and application of university rankings (Gadd et al., 2021; Universities of

the Netherlands (UNL), 2023), including the impossibility to capture the quality of an entire university with different programmes and disciplines into a single digit; the use of selective quantitative data and self-selecting questionnaire responses, with the data themselves not being openly available; and the emphasis on scoring and competition, as opposed to collaboration. Nonetheless, universities are still participating in supplying information to the often commercial organisations producing rankings and publicising their position on these rankings on their website.

There have been calls for change, though. Recently, initiatives like More Than Our Rank (https://inorms.net/more-than-our-rank/) have been developed in response to some of the problematic features and effects of the global university rankings, providing 'an opportunity for academic institutions to highlight the many and various ways they serve the world that are not reflected in their ranking position'. Importantly, the EUA, with over 800 member institutions, explicitly supports More Than Our Rank (EUA, 2022). Although changing rankings usage in reality by individual institutions may prove hard, actual steps to collaboratively approach the issue are being taken (Upton, 2023). Also, at least for researcher assessment, the Coalition for Advancing Research Assessment (COARA, see below) calls to not use university rankings for research assessment, especially because the criteria on which they are built tend to trickle down from the university to the researcher level (COARA, 2022). On the issue of transparency, the Leiden Ranking has announced an initiative to launch a version of their ranking fully based on open data sources (van Eck et al., 2023).

International Developments

In the previous sections, a number of developments in the Netherlands as well as international developments have been mentioned that align and support changes in academic assessment. As alluded to above, it's important that there is both broad support for such changes and that changes are driven by, or developed in close consultation with, researchers and research organisations, rather than imposed on them. Although the Netherlands has in many respects been in the forefront of research assessment reform, in particular regarding national collaboration of stakeholders, that reform has roots in many countries and is supported by many international and regional initiatives as shown by the Future of Research Evaluation report (De Rijcke et al., 2023). Reforms are not only discussed in academia but also in mainstream science media (e.g., Pain, 2023). A selection of four developments, that all provide the opportunity for direct action at the level of research organisations, is elaborated on below as a jump off point for research organisations interested in changing assessment culture at their institution and beyond.

DORA

One of the first initiatives to challenge the widespread use of journal metrics like JIF for assessment of research and researchers has been the San Francisco

Declaration of Research Assessment (DORA) (DORA, 2013), which also suggests alternative approaches for journals, funders and institutions. In the 10 years since DORA was published in 2013, it has collected signatories of 20,479 individuals and 2,866 organisations in 161 countries. It has been a starting point for wide-ranging discussion and explorations on changes in assessment and evaluation. Signing DORA is often an early step in the process of reconsidering research evaluation at a university and at the same time signals commitment towards that process.

COARA

COARA (https://coara.eu/) was started in 2022 with the drafting of the agreement on reforming research assessment (COARA, 2022) by a team of representatives from the EUA, Science Europe and the European Commission, with involvement from more than 350 organisations from over 40 countries were involved. In 2023, COARA opened for signatories worldwide and proceeded to invite proposals for working groups to collectively work on reforming research assessment at European research performing and research funding organisations. As of 26 June 2023, there are 510 COARA member organisations from across the world. Like with DORA, signing COARA signals commitment to changing research assessment, and moreover, COARA can provide opportunities for mutual learning and exchanging best practices among participating institutions.

More Than Our Rank

The More Than Our Rank initiative (https://inorms.net/more-than-our-rank/) was started in 2023 by the International Network of Research Management Societies (INORMS) to provide an opportunity for academic institutions to highlight the many and various ways they serve the world that are not reflected in their ranking position. Academic institutions are asked to post a statement promoting institutional activities, achievements or ambitions that are not adequately captured by national or international university rankings on their website, alongside the More Than Our Rank logo, to expand the information conveyed just by their position on international rankings. This both encourages universities to consider their strengths beyond the usual narrow indicators and makes this visible for all to see.

HELIOS

HELIOS (Higher Education Leadership Initiative for Open Scholarship, https://www.heliosopen.org/) is an initiative in the United States that constitutes a coordinated effort to align higher education practices with open scholarship values, with about hundred institutions committing to it at the time of writing. It addresses values, practices and incentives. In its joint statement on Reforming Hiring, Reappointment, Tenure and Promotion (HELIOS, 2022), it asks institutions to commit to an internal dialogue on having hiring, tenure and promotion

better reflect open research and scholarship. That dialogue should be with all institutional stakeholders, it should seek endorsements from decision-making bodies and include the development of frameworks and guidance. While in the current phase, much is left to the individual institutions, HELIOS does commit to yearly progress updates on these reform strategies.

Conclusion and Discussion

Traditionally, assessment in academia has been focused on individual research performance and within that, on (journal) publications as measurable output, with importance given to quantity and a small number of standardised metrics. Assessment (for hiring, tenure and promotion, grant allocation, prizes, etc.) is usually summative, selecting individuals on the basis of demonstrated achievements. In terms of talent management, this approach to assessment corresponds to an exclusive, subject-oriented approach. The ambition for individuals in the system, as well as for institutions as a whole, is to strive for 'excellence' – in turn often defined by a narrow focus on research outputs and a limited number of indicators.

There is, however, a tension between these views on assessment and talent management and a more holistic view on the role of academic institutions in accelerating and improving science and scholarship and its societal impact. Open science practices as well as concerns around research integrity are challenging the current incentive system. Increasingly, institutions, also with the push from funders and the open science and responsible research movements, are looking to broaden and deepen their assessment practices. This also has implications for how talent management is conceptualised and implemented.

Broadly speaking, three developments can be observed:

- A move from research-focused assessment to assessment including all academic activities (including education, professional performance and leadership).
- A shift from focus on the individual to a focus on collaboration in teams, where individuals can have different roles, and contributions of both research staff and support staff are recognised as important in reaching the group's goals.
- Critical reflection on any indicators used – making sure they are both fit for purpose (i.e., whether what they measure indeed reflects the qualities that the evaluation is intended to assess) and fit for the level at which assessment is taking place.

For talent management, these developments reflect a shift from exclusive, subject-oriented talent management (with the aim of selecting 'the best' individuals in isolation) to inclusive, object-oriented talent management (with an eye for the qualities, expertises and competences needed at the team level to reach its strategic goals, which in themselves can be broader and more varied).

An aspect that deserves special attention in implementation of this view on talent management is agency. First of all, this means agency of the subjects of assessment (be it research groups and the individuals within those groups, or

universities as a whole), who ideally should have a say in the strategic goals, activities and outputs they are assessed on, as well as involved in the design of the assessment process. In addition, it involves agency in control of the data, tools and platforms that are used for assessment and talent management, to ensure assessment is not limited a priori to data and methods chosen by (often commercial) providers of these tools and platforms, as this effectively amounts to outsourcing part of talent management to external organisations with their own commercial interests (including providers of bibliographic databases, research analytics tools and university rankings).

Implementing such changes in assessment in academia, including in talent management, requires involvement from university management, human resource management and academic and support staff at all career levels. Small steps can be taken in signing relevant declarations (such as DORA) as a publicly visible signal of commitment, participating in relatively low-risk but meaningful initiatives like More Than Our Rank and contributing to and benefitting from mutual learning exercises as organised through COARA and HELIOS to discuss approaches and good practices. As more institutions take steps on the road towards more inclusive, relevant and responsible assessment and talent management, the easier it becomes for other institutions to follow the same path, always allowing for local and disciplinary contexts.

References

Advisory Council for Science, Technology and Innovation (AWTI). (2023). *Assessing the qualities of science*. https://www.awti.nl/binaries/awti/documenten/publicaties/2023/06/13/assessing-the-qualities-of-science/Assessing±the±qualities±of±science.pdf

Aguinis, H., Cummings, C., Ravini, R., & Cummings, T. G. (2020). "An A is an A": The new bottom line for valuing academic research. *Academy of Management Perspectives*, *34*, 135–154. https://doi.org/10.5465/amp.2017.0193

Aubert Bonn, N., & Bouter, L. (2023). Research assessments should recognize responsible research practices. Narrative review of a lively debate and promising developments. In E. Valdés & J. A. Lecaros (Eds.), *Handbook of bioethical decisions. Volume II. Collaborative bioethics* (Vol. 3, pp. 441–472). Springer. https://doi.org/10.1007/978-3-031-29455-6_27

Bendiscioli, S. (2018). The troubles with peer review for allocating research funding. *EMBO Reports*, *20*, e49472. https://doi.org/10.15252/embr.201949472

Bertocchi, G., Gambardella, A., Jappelli, T., Nappi, C. A., & Peracchi, F. (2015). Bibliometric evaluation vs. informed peer review: Evidence from Italy. *Research Policy*, *44*, 451–466. https://doi.org/10.1016/j.respol.2014.08.004

Bordignon, F., Chaignon, L., & Egret, D. (2023). Promoting narrative CVs to improve research evaluation? A review of opinion pieces and experiments. *Research Evaluation*, *32*(2), 313–320. https://doi.org/10.1093/reseval/rvad013

Bornmann, L., & Daniel, H.-D. (2009). The state of *h* index research. Is the *h* index the ideal way to measure research performance? *EMBO Reports*, *10*(1), 2–6. https://doi.org/10.1038/embor.2008.233

Boselie, P. (2014). Performance management. In *Strategic human resource management: A balanced approach* (2nd ed., pp. 169–188). McGraw Hill.

Bouter, L. (2023). Why research integrity matters and how it can be improved. *Accountability in Research*, *9*, 1–10. https://doi.org/10.1080/08989621.2023.2189010

COARA. (2022). *Agreement on reforming research assessment*. https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf

Cole, S., Cole, J. R., & Simon, G. A. (1981). Chance and consensus in peer review. *Science*, *214*(4523), 881–886. https://doi.org/10.1126/science.7302566

Curry, S. (2018, June 12). *I am not my h-index (or my JIFs)*. [Image attached] [Tweet]. Twitter. https://twitter.com/Stephen_Curry/status/1006539579426508802

De Rijcke, S., Waltman, L., & van Leeuwen, T. (2021, May 19). *Halt the* h-*index*. Leiden Madtrics. https://www.leidenmadtrics.nl/articles/halt-the-h-index

De Rijcke, S., Cosentino, C., Crewe, R., Motala-Timol, S., Binti a Rahman, N., Rovelli, L., Vaux, D., & Yupeng, Y. (2023). *The future of research evaluation: a synthesis of current debates and developments* [Discussion paper]. https://doi.org/10.24948/2023.06

DORA. (2013). *San Francisco declaration on research assessment*. Centre for Science Futures https://sfdora.org/read/

Dutch Research Council (NWO). (n.d.). *DORA*. Retrieved July 9, 2023, from https://www.nwo.nl/en/dora

European University Association (EUA). (2022). *More than our rank: EUA supports global initiative on responsible assessment systems*. https://eua.eu/news/982:more-than-our-rank-eua-supports-global-initiative-on-responsible-assessment-systems.html

Gadd, E. (2019, May 15). *The blind and the elephant: Bringing clarity to our conversations about responsible metrics*. The Bibliomagician. https://thebibliomagician.wordpress.com/2019/05/15/the-blind-and-the-elephant-bringing-clarity-to-our-conversations-about-responsible-metrics/

Gadd, E., Holmes, R., & Shearer, J. (2021). Developing a method for evaluating global university rankings. *Scholarly Assessment Reports*, *3*(1), 1–19. https://doi.org/10.29024/sar.31

Gopalakrishna, G., Ter Riet, G., Vink, G., Stoop, I., Wicherts, J. M., & Bouter, L. M. (2022). Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in The Netherlands. *PloS One*, *17*(2), e0263023. https://doi.org/10.1371/journal.pone.0263023.

Gopalakrishna, G., Wicherts, J. M., Vink, G., Stoop, I., van den Akker, O, Ter Riet, G., & L. Bouter, L. (2022). Prevalence of responsible research practices among academics in the Netherlands [version 2; peer review: 2 approved]. *F1000Research*, *11*, 471. https://doi.org/10.12688/f1000research.110664.2

Gossink-Melenhorst, K. (2019, November 14). *Quality over quantity: How the Dutch Research Council is giving researchers the opportunity to showcase diverse types of talent*. DORA. https://sfdora.org/2019/11/14/quality-over-quantity-how-the-dutch-research-council-is-giving-researchers-the-opportunity-to-showcase-diverse-types-of-talent/

HELIOS. (2022). *Joint statement on reforming hiring, reappointment, tenure and promotion*.

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden manifesto for research metrics. *Nature*, *520*(7548), 429–431. https://doi.org/10.1038/520429a

Huang, C.-K., Neylon, C., Montgomery, L., Hosking, R., Diprose, J. P., Handcock, R. N., & Wilson, K. (2022). *Open access research outputs receive more diverse citations (version 2)*. Zenodo. https://doi.org/10.5281/zenodo.7099438

Kallio, K.-M., Kallio, T. J., & Grossi, G. (2017). Performance measurement in universities: Ambiguities in the use of quality versus quantity in performance indicators. *Public Money & Management*, *37*, 293–300. https://doi.org/10.1080/09540962.2017.1295735

Kaltenbrunner, W., Haven, T., Algra, A., Akse, R., Arici, F., Bakk, Z., Bandola, J., Chan, T. T., Costas Comesana, R., Coopmans, C., Csiszar, A., de Bordes, C., Dudek, J., Etheridge, M., Gossink-Melenhorst, K., Hamann, J., Hammarfelt, B., Hoffmann,

M., Ipema, Z., … Wouters, P. (2023, March 25). *Narrative CVs: A new challenge and research agenda*. Leiden Madtrics. https://www.leidenmadtrics.nl/articles/narrative-cvs-a-new-challenge-and-research-agenda

Larivière, V., Kiermer, V., MacCallum, C. J., McNutt, M., Patterson, M., Pulverer, B., Swaminathan, S., Taylor, S., & Curry, S. (2016). *A simple proposal for the publication of journal citation distributions*. BioRxiv. https://doi.org/10.1101/062109

Larivière, V., & Sugimoto, C. R. (2019). The journal impact factor: A brief history, critique, and discussion of adverse effects. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Handbook of science and technology indicators* (pp. 3–24). Springer. https://doi.org/10.1007/978-3-030-02511-3_1

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, *64*(1), 2–17. https://doi.org/10.1002/asi.22784

McKiernan, E. C., Schimanski, L. A., Muñoz Nieves, C., Matthias, L., Niles, M. T., & Alperin, J. P. (2019). Use of the journal impact factor in academic review, promotion, and tenure evaluations. *Elife*, *8*, e47338. https://doi.org/10.7554/eLife.47338

Moher, D., Naudet, F., Cristea, I. A., Miedema, F., Ioannidis, J. P., & Goodman, S. N. (2018). Assessing scientists for hiring, promotion, and tenure. *PLoS Biology*, *16*(3), e2004089. https://doi.org/10.1371/journal.pbio.2004089

Moore, S., Neylon, C., Eve, M. P., O'Donnell, D. P., & Pattinson, D. (2017). "Excellence R Us": University research and the fetishisation of excellence. *Palgrave Communications*, *3*(1), 1–13. https://doi.org/10.1057/palcomms.2016.105

Pain, E. (2023). *How academia is exploring new approaches for evaluating researchers*. Science. https://doi.org/10.1126/science.caredit.adj4869

Plomp, E., Schettino, A., & Tsang, E. (2021, September 1). *Quality is key*. Open Working. https://openworking.wordpress.com/2021/09/01/quality-is-key/

Saenen, B., & Borrell-Damián, L. (2019). *Reflections on university research assessment. Key concepts, issues and actors: EUA briefing*. https://eua.eu/component/attachments/attachments.html?id=3446

Singh Chawla, D. (2019). *Thousands of grant peer reviewers share concerns in global survey*. Nature News. https://doi.org/10.1038/d41586-019-03105-2

Teplitskiy, M., Acuna, D., Elamrani-Raoult, A., Körding, K., & Evans, J. (2018). The sociology of scientific validity: How professional networks shape judgement in peer review. *Research Policy*, *47*(9), 1825–1841. https://doi.org/10.1016/j.respol.2018.06.014

Thunnissen, M., Boselie, P., & Fruytier, B. (2013). A review of talent management: 'Infancy or adolescence?'. *The International Journal of Human Resource Management*, *24*(9), 1744–1761. https://doi.org/10.1080/09585192.2013.777543

Thunnissen, M., van Arensbergen, P., & van den Brink, M. (2021). Talent management in academia. In I. Tarique (Ed.), *The Routledge companion to talent management* (pp. 215–226). Routledge. https://doi.org/10.4324/9781315474687

Traag, V. A., & Waltman, L. (2019). Systematic analysis of agreement between metrics and peer review in the UK REF. *Palgrave Communications*, *5*, 29. https://doi.org/10.1057/s41599-019-0233-x

UNESCO. (2021). *UNESCO recommendation on open science*. UNESCO. https://doi.org/10.54677/MNMH8546

Universities of the Netherlands (UNL). (2023). *Ranking the university: On the effects of rankings on the academic community and how to overcome them*. https://universiteitenvannederland.nl/files/documents/Publications/Ranking_the_university_ENG.pdf

Universities UK. (2019). *The concordat to support research integrity*. https://www.universitiesuk.ac.uk/sites/default/files/field/downloads/2021-08/Updated%20FINAL-the-concordat-to-support-research-integrity.pdf

University Medical Centre Utrecht (UMCU). (2022). *Implementing new recognition & rewards through six new academic career profiles* [video]. YouTube. https://youtu.be/iRBCXUC3NDE

University Medical Centre Utrecht (UMCU). (2023). Career Profiles: Researchers [presentation]. https://assets-eu-01.kc-usercontent.com/546dd520-97db-01b7-154d-79bb6d950a2d/5b97a599-5b3c-4727-a995-bc07038dae0b/UMCU%20Academic%20Career%20Profiles_May2023.pdf

Upton, B. (2023). *European rectors rally around joint position on rankings*. Times Higher Education. https://www.timeshighereducation.com/news/european-rectors-rally-around-joint-position-rankings

Utrecht University. (2021). *Utrecht University recognition and rewards vision*. https://www.uu.nl/sites/default/files/UU-Recognition-and-Rewards-Vision.pdf

van den Brink, M., & Benschop, Y. (2014). Gender in academic networking: The role of gatekeepers in professorial recruitment. *Journal of Management Studies*, *51*(3), 460–492. https://doi.org/10.1111/joms.12060

van Eck, N. J., Waltman, L., Anli, Z., Calero-Medina, C., Gibson, D., Neijssel, M., Reyes Elizondo, A., Visser, M., de Winter, H., & Yegros, A. (2023, June 21). *The CWTS Leiden ranking 2023*. Leiden Madtrics. https://www.leidenmadtrics.nl/articles/the-cwts-leiden-ranking-2023

Vereniging van Universiteiten (VSNU). (2018). *Netherlands code of conduct for research integrity*. Royal Netherlands Academy of Arts and Sciences (KNAW), Netherlands Federation of University Medical Centers (NFU), Dutch Research Council (NWO), TO2 Federations, Netherlands Association of Universities of Applied Sciences (VH), Vereniging van Universiteiten (VSNU, now UNL). https://www.universiteitenvannederland.nl/files/documents/Netherlands%20Code%20of%20Conduct%20for%20Research%20Integrity%202018.pdf

Vereniging van Universiteiten (VSNU). (2019). *Room for everyone's talent*. Vereniging van Universiteiten (VSNU, now UNL), Netherlands Federation of University Medical Centers (NFU), Royal Netherlands Academy of Arts and Sciences (KNAW), Dutch Research Council (NWO), Netherlands Organisation for Health Research and Development (ZonMw). https://www.universiteitenvannederland.nl/recognitionandrewards/wp-content/uploads/2019/11/Position-paper-Room-for-everyone%E2%80%99s-talent.pdf

Vereniging van Universiteiten (VSNU). (2020). *Strategy evaluation protocol (2021-2027)*. Vereniging van Universiteiten (VSNU, now UNL), Royal Netherlands Academy of Arts and Sciences (KNAW), Dutch Research Council (NWO). https://www.universiteitenvannederland.nl/files/documenten/Domeinen/Onderzoek/SEP_2021-2027.pdf

Woolston, C. (2022). Time to rethink the scientific CV. *Nature*, *604*(7904), 203–205. https://doi.org/10.1038/d41586-022-00928-4