Research Assessment:
Recognising the asset of diversity for scholarship serving society

Martin Dominik

University of St Andrews, Centre for Exoplanet Science, SUPA School of Physics & Astronomy, North Haugh, St Andrews, KY16 9SS, United Kingdom
(md35@st-andrews.ac.uk)

Abstract

Diversity is one of humanity's greatest strengths, and not a competing goal to excellence, right to the contrary. Moreover, the quality of teams hinges on the complementary skills and interaction between its members. However, metrics for assessing research that have been popular over the recent two decades critically fail on these fundamental principles for advancing science to the benefit of society. Research management has become obsessed with outputs, productivity, and citations; none of which reflect good science. University rankings must not serve as excuse for bad management, and research assessment must never become disconnected from a specific purpose. Institutions need to set their own policies that let them thrive within their niche in the global research ecosystem rather than trying to engage in meaningless overcompetition. I will suggest a simple framework for fostering an environment that supports creativity and initiative.

The omnipresence of evaluation and the absence of judgement

Evaluation has become omnipresent. You cannot travel anymore without being asked to rate your flight, taxi driver, hotel, or restaurant, and once back home your opinion about the cashier in your local supermarket is requested. It should therefore not surprise that the rise of scientometrics over the last 15 years[1] follows this fashion trend. Notably, these are not rooted in tradition, the h-index specifically was only invented in 2005.[2] Fashions come, and they go. More recently, a push for quantifying so-called "impact" means that there now is hardly any public engagement with science without feedback form.

This obsession with metrics comes along with a striking absence of sensible judgement in management. As Jerry Muller in his book "The tyranny of metrics" stresses,[3] "Measurement is not an alternative to judgement. Measurement rather demands judgement:

- judgement about *whether* to measure,
- *what* to measure,
- how to evaluate the *significance* of what's been measured,
- whether *rewards and penalties* will be attached to the results,
- and *to whom* to make the measurements available."

Simply delegating decisions to numbers means dodging responsibility.

The business of productivity and popularity metrics

Metrics have become a much profitable business, exploiting the prevalent lack of judgement. Large publishers have turned into IT giants claiming to provide information-based analytics and decision tools. But those "tools" have turned into instruments of power. Once you control publishing, and

[1] e.g., J. Wilsdon: "*The Metric Tide: Independent Review of the Role of Metrics in Research Assessment and Management*", SAGE Publications Ltd. DOI: 10.4135/9781473978782 (2015)

[2] J. E. Hirsch: "*An index to quantify an individual's research output*", PNAS 102, 16569. DOI: 10.1073/pnas.0507655102 (2005)

[3] J. Z. Muller: "*The tyranny of metrics*", Princeton University Press. ISBN 9780691174952 (2018)

you control assessment metrics, and you can claim proprietary rights to the underlying data, you not only take ownership of the research process, but by shaping your gigantic scientific industrial complex you also redefine societal values. – We need to break free from that.

"Productivity" is a much ill-suited concept when trying to assess the research ecosystem. Paul Smaldino and Richard McElreath argued that incentives for publication quantity can drive the evolution of poor methodological practices through what they called "The natural selection of bad science".[4] The worse your methods, the more false positives are claimed as exciting discoveries. The more of those you produce, the more funds, the more students, – and the worse the science, the higher its replication rate. How do we break that replication cycle?

Citation counts and derived metrics tell us what is popular. They are not objective, but result from subjective decisions, providing us with a map of social networks. That is, social processes do not create a bias to citation counts, but these are what citations are measuring in the first place. Judging by network completely contravenes recognising those who are marginalised. The brightest kid in school did not tend to be the most popular, and the scientific mainstream is at least a decade behind the forefront of innovation. Those who are much ahead do not have many followers, because they think unlike the majority.

Diversity as a key asset for advancing knowledge for the benefit of society

So, how do we move forward towards building an efficient research ecosystem that advances knowledge for the benefit of society? We need a trustworthy public record of science, part of shared human culture. Advances need to be communicated so that others are able to build on these, eventually leading to translation into societal benefits. There is no dichotomy between fundamental and applied research, that is the wrong fight to take. All research that significantly adds to the scholarly record holds the potential of being translated into concrete value for society be it sooner or be it later, not necessarily by those who originally carried out that research.

Many approaches to research assessment are based on the fundamental misconception of the universality of research "excellence". Best for what? Meaningful assessment needs to serve a specific purpose, and there is no meaningful universal ranking, neither at global, national, regional, or institutional level. Depending on specific context, research assessment is to take various forms und is to use various criteria. Assessment metrics tell us that the answer is 42, but we have forgotten what the question is.[5]

Biodiversity is a key feature of biological evolution, and similarly the research ecosystem is and should be diverse at various levels of granularity. In fact, diversifying to occupy every niche in an ecosystem provides a much larger value than engaging in a competition that has one winner only, and diversity provides resilience.[6]

Team roles, track records, and intrinsic skills

Actors within a system do not operate in isolation, but they interact within specific context. This holds at every level and scale, from research group to global society. Diversity again emerges as one

---

[4] P. E. Smaldino & R. McElreath: "*The natural selection of bad science*", R. Soc. Open Sci. 3, 160384.
   DOI: 10.1098/rsos.160384 (2016)

[5] c.f., D. Adams: "*The Hitchhiker's Guide to the Galaxy*", Arthur Barker Limited, London. ISBN 9780213167387 (1979)

[6] e.g., A. Kershenbaum: "*The Zoologist's Guide to the Galaxy: What Animals on Earth Reveal about Aliens – and Ourselves*", Penguin Books. ISBN 9781984881984 (2020)

of our greatest strengths, making our society thrive and letting us work together. It is not a competing goal to excellence, right to the contrary.

The quality of teams hinges on the complementarity of skills and on the interaction between its members. Those who are quite similar and interchangeable do not make a good team. Relevant assessment criteria do not trickle down from larger units to smaller ones: you do not form a good football team from those players who scored the largest number of goals. People successfully use their skills in various roles, and practices that are relevant in one role might not be relevant in another.

However, the freedom of diversity of practices should belong to each individual researcher, not to so-called "disciplines". We need an overarching framework that appreciates diversity of individuals and does not lock them up in "disciplines".

Expecting specific track records is a key obstacle to mobility and diversity of career paths, neglects specific circumstances, and ultimately is a great disservice both to those assessed and to society. In particular, there are substantial differences between academic and industrial environments, governmental and not-for-profit institutions. If something was not relevant in a previous role, there is no track record of it. Measures of productivity again fail on the diversity of environments, conditions, and circumstances. While the central point of the San Francisco Declaration on Research Assessment (DORA) is not to judge publications by where they have appeared,[7] we also might want to consider not judging researchers by where they have appeared. We need to distil and value their intrinsic traits, skills, and competencies.

Supporting initiative and the integrity of the research process

If we look at research and its benefits, we should carefully distinguish between processes, outputs, and outcomes. Specifically, outputs is what you produce, outcomes is the difference it makes. While it has been most popular to judge the research process by its outputs, more recently the focus has shifted to also consider judging the research process by its outcomes. But I would argue that we should rather judge outputs and outcomes by the process, valuing the journey rather than the destination reached. We should be looking for those who are good at practice, not those who happened to be lucky. Rewarding luck is not any better than choosing at random. In fact, the Hong Kong Principles, developed during the 6th World Conference on Research Integrity in 2019,[8] place the scientific process and good scholarly citizenship at their core, emerging from the fact that "knowledge must be trustworthy to benefit research and society", and "trustworthy research is robust, rigorous, and transparent at all stages of design, execution and reporting".

We are discussing peer review in the context of scholarly publishing, but it actually is a form of research assessment. How should we develop it and make good use of it?

A specific quality of the scientific process is its efficiency and the economical usage of resources. Some prevalent assessment metrics right to the contrary celebrate wasting money. I would be quite interested to see US Ivy League universities compared to Central African universities on this point.

In the policy sphere, I keep hearing about how to incentivise, but we should talk more about supporting initiative and building on human creativity. I think that the better strategies are built *with* people, rather than *for* them. You definitely do not foster leadership by already predetermining the direction of the journey. With creative freedom comes responsibility, and our societal

[7] DORA: "*Declaration on Research Assessment*". https://sfdora.org/.

[8] D. Moher et al.: "*The Hong Kong Principles for assessing researchers: Fostering research integrity*". PLoS Biology 18, e3000737. DOI: 10.1371/journal.pbio.3000737 (2020)

development requires empowered and responsible actors, able to decide for themselves what is right or wrong. Ethical behaviour stems from responsibility, not from rulebooks.

If assessment however operates within fixed expectations, there is no room for acknowledging outstanding contributions; this requires flexible goalposts. High-risk research strategies have their place in a well-mixed portfolio, and it would be disingenuous to blame individuals for their failure.

Five pillars of purposeful research assessment

So, to summarise, purposeful research assessment can be built on 5 major pillars:

- Specificity and purpose,
- diversity of practice and team roles,
- intrinsic traits and diversity of career paths,
- qualities of research processes, as well as
- creativity and initiative.

Diversity is our key asset, and actors need the freedom to be "different".

Change and outer pressure

There is outer pressure, however. Notably, signing up to the idea that university rankings confer valuable prestige, institutions enter a competition in which almost everyone will be a loser. Outsourcing values to such rankings[9] stands in the way of meaningful development that lets institutions flourish in their specific niche that sits within a diverse system. Some of these rankings might be interesting, but attaching undue relevance to them is fatal. They must not serve as excuse for bad management. Critical judgment is required.

If you want to see changes happening, be assured that you are not alone. But that change needs you, do not hesitate to call out nonsense for what it is and join the efforts, – *you* will make the difference.

[9] e.g., E. Gadd: *"Mis-Measuring Our Universities: Why Global University Rankings Don't Add Up"*, Frontiers in Research Metrics and Analytics. DOI: 10.3389/frma.2021.680023 (2021)

[10] M. Dominik et al.: "*Publication models, assessment, and open science*", Global Young Academy. https://globalyoungacademy.net/wp-content/uploads/2018/10/APOS-Report-29.10.2018.pdf (2018)

Pillars of Purposeful Research Assessment

Specificity and purpose

Research assessment must never be disconnected from an underlying specific purpose and needs to address a concrete question rather than being based on a universal (mis-)concept of research "excellence". Depending on specific context, research assessment can take various forms and use various criteria.

Diversity of practices and team roles

Respect that the scientific ecosystem thrives from a wide diversity of practices and team roles, evidenced by a broad range of outputs and activities. Recognise that a multitude of such practices and roles might apply to an individual researcher, and that these keep evolving.

Intrinsic traits and diversity of career paths

Value the intrinsic traits, skills, and competencies of researchers rather than conflating assessment with their career path. Recognise that researchers operate in various environments, under various conditions and circumstances.

Qualities of research processes

View outputs and outcomes on how they are rooted in and how they are reflective of research processes. Explicitly verify that research processes are rigorous and adhere to good standards of integrity and ethics.

Creativity and initiative

Support the creativity and initiative of researchers, including the pursuit of original and/or high-risk strategies. Rewarding outstanding contributions requires flexible goalposts.