

# World view



By Anna Hatch

## To fix research assessment, swap slogans for definitions

**Evaluation reforms will go round in circles without conceptual clarity, warns Anna Hatch.**

**T**he need for clarity extends beyond how we communicate science to how we evaluate it. Who can really define stock phrases such as ‘a significant contribution to research’? Or understand what ‘high impact’ or ‘world-class’ mean?

Seven years ago this month, scientists met in San Francisco, California, to call for an end to the practice of assessing research through the impact factors of the journals in which it is published. They demanded that institutions instead be explicit about their criteria and consider all scholarly outputs – preprints, code, data, peer review, teaching, mentoring and so on. Today, thousands have signed the resulting Declaration on Research Assessment (DORA). But actual change is all too slow.

Two years ago, the DORA steering committee hired me to survey practices in research assessment and promote the best ones. Other efforts have similar goals. These include the Leiden Manifesto and the HuMetricsHSS Initiative.

My view is that most assessment guidelines permit sliding standards: instead of clearly defined terms, they give us feel-good slogans that lack any fixed meaning. Facing the problem will get us much of the way towards a solution.

Broad language increases room for misinterpretation. ‘High impact’ can be code for where research is published. Or it can mean the effect that research has had on its field, or on society locally or globally – often very different things. Yet confusion is the least of the problems. Descriptors such as ‘world-class’ and ‘excellent’ allow assessors to vary comparisons depending on whose work they are assessing. Academia cannot be a meritocracy if standards change depending on whom we are evaluating. Unconscious bias associated with factors such as a researcher’s gender, ethnic origin and social background helps to perpetuate the status quo. It was only with double-blind review of research proposals that women finally got fair access to the Hubble Space Telescope. Research suggests that using words such as ‘excellence’ in the criteria for grants, awards and promotion can contribute to hypercompetition, in part through the ‘Matthew effect’, in which recognition and resources flow mainly to those who have already received them.

Many strategies exist to improve equity in academia, but conceptual clarity is paramount. A study probing the use of ‘outcome’ and ‘impact’ in international-development work concluded that such terms undermine evaluation efforts. It proposed a combination of strategies including the use of meaningful qualifiers, such as the type of result and how it relates to a project’s purpose, and the creation of mutually exclusive definitions for terms such as ‘outcome’,

  
**Most assessment guidelines permit sliding standards.”**

‘impact’ and ‘output’ (B. Belcher & M. Palenberg *Am. J. Eval.* 39, 478–495; 2018).

Some people say that excellence is easy to identify because ‘you know it when you see it’. But Nobel prizes have been awarded for research that was not immediately recognized as a major breakthrough. And it becomes practically impossible to distinguish shades of excellence when many qualified applicants compete for limited funds.

Being explicit about how specific qualities are valued leads assessors to think critically about whether those qualities are truly being considered. Achieving that conceptual clarity requires discussion with faculty members, staff and students: hours and hours of it. The University Medical Center Utrecht in the Netherlands, for example, held a series of conversations, each involving 20–60 researchers, and then spent another year revising its research assessment policies to recognize societal impacts.

Although DORA curates examples of good practice (see [go.nature.com/2qkcssw](http://go.nature.com/2qkcssw)), most of the best efforts cannot (yet) be found in databases or publications. Often the only way to learn about them is through discussion and networking. It was not until DORA held a meeting with the Howard Hughes Medical Institute in Chevy Chase, Maryland, in October that I learnt the University of California, Irvine, had moved to include collaborative scholarship in evaluations. It took an e-mail exchange to learn of an administrator’s personal efforts to find tools that explicitly credit collaboration.

Frank conversations about what is valued in a particular context, or at a specific institution, are an essential first step in developing concrete recommendations. Although ambiguous terms, for instance ‘world-class’ and ‘significant’, are a hindrance when performing assessments, university administrators have also told me that they rely on flexible language to make room to reward a variety of contributions. So it makes sense that more specific language in review, promotion and tenure guidelines must be able to accommodate varied outputs, outcomes and impacts of scholarly work.

The joint meeting of the American Society for Cell Biology and the European Molecular Biology Organization in Washington DC this month will include a mock faculty-recruitment exercise, involving approaches such as removing applicant names and journal titles from bibliographies. Participants will then discuss which standards to apply to improve objectivity, and how to apply them.

Setting such standards will be tough. It will be tempting to fall back on the misleading simplicity of metrics such as impact factors, or on ambiguous terms that can be agreed to by everyone but applied judiciously by no one. It is too early to know what those standards will be or how much they will vary, but the right discussions are starting to happen. They must continue.

**Anna Hatch** is programme director for the Declaration on Research Assessment in Bethesda, Maryland. e-mail: [ahatch@sfordora.org](mailto:ahatch@sfordora.org)