

Research evaluation needs to change with the times

The focus on a narrow set of metrics leads to a lack of diversity in the types of leader and institution that win funding.

Many researchers who are funded from public sources are required to participate in national evaluations of their work. Such assessments are popular with governments because they help to ensure a degree of accountability for taxpayer cash. Funders like them, too, because they provide a useful benchmark for the standard of research being done. Universities also benefit financially when they write their research strategies around the requirements of assessments. By contrast, researchers generally see assessments as unhelpful to their work. Evaluations can also be stressful and burdensome, and in some cases create tensions between colleagues in academic and administrative roles.

With a few exceptions, the principal components of assessment systems have stayed largely the same since the exercises began, in the 1980s. But some countries are contemplating reworking these systems to reflect how science is done today. Change has been a long time coming, precipitated by initiatives such as the 2013 San Francisco Declaration on Research Assessment, the 2015 Leiden Manifesto for research metrics and the 2020 Hong Kong Principles for assessing researchers. Official research assessments are clearly behind the times and need to catch up.

Last November, the European Commission announced plans to put together a European Union-wide agreement on research assessment. It is proposing that assessment criteria reward ethics and integrity, teamwork and a diversity of outputs in addition to research quality and impact. The UK Future Research Assessment Programme, due to report by the end of this year, has also been tasked with proposing ways to ensure that assessments become more inclusive. These changes cannot come soon enough.

Measures of success

Research-assessment systems are the nearest thing that universities have to the performance metrics that are common in business. Individual researchers are assessed on a range of measures, such as the number and quality of journal articles, books and monographs they have published; their research income; the number of their students who complete postgraduate degrees; and any non-academic impact from their work, such as its influence on society or policy. In the United Kingdom, for example, this information is compressed into a composite

index and the results are used to allocate funding.

UK public funding goes preferentially to the university departments with the highest-performing researchers. But assessments that measure individual performance make it harder for institutions to recognize science conducted in teams – both within and between disciplines. Moreover, research assessments have tended to focus on final published results, whereas researchers are increasingly producing more diverse outputs, including data sets, reproducibility studies and registered reports, in which researchers publish study designs before starting experiments. Most current assessments do not value mentorship and struggle to recognize the needs of researchers from minority communities.

And then there's the question of costs. The 2014 iteration of the UK Research Excellence Framework – the exercise takes place roughly every seven years – cost somewhere in the region of £246 million (US\$334 million). The lion's share (£232 million) was borne by universities. It included the costs of academic staff who served on the review panels that assessed around 190,000 outputs in 36 subject areas; and the costs to institutions, which go to great lengths to prepare their staff, including running mock assessment exercises. Here, smaller institutions lack the resources to compete with better-funded ones.

Researchers who study assessment methods regularly put forward ideas for how evaluations could change for the better. Last August, a working group from the International Network of Research Management Societies fleshed out a framework called SCOPE. This encourages funders to design evaluation systems around the 'values' they wish to assess. For example, rewarding competitive behaviour might require a different set of criteria from incentivizing collegiality. The SCOPE framework also proposes that funders collaborate with the people being evaluated to design the assessment, and urges them to work with experts in research evaluation – a defined research field.

The importance of co-design cannot be overstated: it will enable the views of different research stakeholders to be represented, and ensure that no single voice dominates. Large, research-intensive institutions often do well in conventional evaluations, because they focus their multi-year strategies on attracting and retaining researchers who meet the criteria of success at publishing results and bringing in income, among other things.

Smaller institutions cannot always compete on these grounds – but could gain if future assessments include new criteria, such as rewarding collaborations, or if assessments put less weight on ability to obtain research funding. A broader range of evaluation criteria could ensure that a greater diversity of institutions have opportunities to do well. And that has to be welcomed.

Larger institutions should not in any way feel threatened by these changes. It is often said – in this journal and elsewhere – that making research culture more welcoming requires systemic change. Research evaluation is core to the research system. If evaluation criteria can be made more representative of how research is done, that much-needed culture change will move one important step closer.

“Making research culture more welcoming requires systemic change.”