

An opportunity to reassess research evaluation

Australia's overhaul of how a huge science funder measures research quality offers a chance to review what makes an assessment system work for everyone.

Research evaluation at the Australian Research Council (ARC), one of the country's main funding bodies, is set to get a makeover. Last month, an independent expert review recommended that the ARC scrap its 13-year-old research-scoring system, known as Excellence in Research for Australia (ERA), and its companion exercise, the Engagement and Impact assessment, which grades the real-world benefits of institutions' work. Both had been on hold since last August, awaiting the findings of the review.

This is a rare case in which an evaluation system can be rewritten from scratch. The ARC should take this opportunity to improve how it measures and communicates the value of Australia's research workforce, on the basis of not just lessons learnt from the ERA's deficiencies, but also principles that have been developed and implemented elsewhere in the world. In doing so, it will help to create a research culture that reflects the best possible values that research should represent.

Between 2010 and 2018, there were four ERA exercises scoring outputs from Australia's 42 universities and publicly funded research institutions. Research is given a rating from 1, for "well below world standard", to 5, for "well above world standard". For most science and engineering disciplines, citation metrics are the primary measure; for humanities, social sciences, computing and mathematics, the reliance is on peer review.

Last month's expert review of the ERA, initiated by the Australian government, concludes that the evaluation has meant Australia performs favourably on international benchmarks. But this has come at a cost. Submissions to the review confirmed that the ERA process was onerous, owing to the time taken to compile the information required. This took a toll on both individuals and institutions. At the University of Sydney, for instance, upwards of 40,000 hours of staff time was spent on the ERA process, costing the university more than Aus\$2 million (US\$1.3 million) in salaries alone, according to its submission.

These efforts came without much discernible reward. The results of the ERA – unlike, say, the Research Excellence Framework in the United Kingdom – are not used to decide university funding. Instead, they are used for such purposes such as setting research strategy, benchmarking and evaluating trends. At the same time, pitting institutions

“Pitting institutions against each other encourages them to poach academics and duplicate expertise.”

against each other encourages them to poach academics and duplicate expertise, rather than fostering a collaborative, cross-disciplinary research culture.

The review wisely resists suggesting that the ERA be replaced with a purely metrics-based exercise to reduce overheads. Using metrics such as citations and journal impact factors is problematic, to say the least. For one thing, they do not capture replication studies and meta-analyses, which journals are increasingly publishing. Furthermore, studies in the social sciences and humanities – more frequently published in national, rather than international publications – are especially disadvantaged by such measures. And just counting citations in well-known journals doesn't do justice to other important work, such as building databases and software, and engaging with the public.

Ten years ago this month, the San Francisco Declaration on Research Assessment (DORA) enshrined the principle that evaluations should not rely too heavily on journal impact factors as a measure of research quality. The declaration, developed during the 2012 Annual Meeting of the American Society for Cell Biology, urges, among other things, that journal-based metrics not be used as a surrogate measure of research quality for assessing an individual scientist's contributions, or in hiring, promotion or funding decisions. More than 20,000 individuals and 2,800 institutions across 160 countries have signed DORA so far.

More has happened in the decade since. The 2015 Leiden Manifesto recommends that assessors anticipate how appraisal systems can be gamed (D. Hicks *et al. Nature* 520, 429–431; 2015). The 2019 Hong Kong Principles aim to build trust in science by including transparency, open-science and responsible-research practices in evaluations (D. Moher *et al. PLoS Biol.* 18, e3000737; 2020). In 2021, the International Network of Research Management Societies developed the SCOPE Framework, a how-to guide for conducting assessments that embed these desirable principles.

Australia is not alone in reconsidering how it does appraisals. The Swedish Research Council, for instance, avoids the pressure of evaluating its entire research workforce in a single all-encompassing exercise by assessing specific disciplines as required, rather than according to a set cycle. The Netherlands, too, has a less onerous process that doesn't just capture research institutions' previous outputs, but looks forwards by identifying future plans and capacity, to pinpoint strengths and areas for improvement ahead of time.

The ERA's replacement should also not downgrade the expertise of research managers. Many are researchers themselves, and their role extends far beyond administering the ERA. Through their work, universities are continually striving to improve policies surrounding research ethics, public engagement and publishing, seeking to become more diverse and more welcoming places. It would be folly to lose that expertise.

It's important to score the quality of publicly funded research. There's no perfect system, but the ARC and other national funding agencies should take care to adopt principles and practices that work. Evaluation has the power to shape research culture for good and ill. It needs to be done well.