



DALE EDWIN MURRAY

RESEARCH ASSESSMENT

The limits of excellence

Young researchers and interdisciplinary science might be getting short-changed by research assessment in Australia and New Zealand.

BY ANNABEL MCGILVRAY

The day before he and I speak, Jonathan Boston receives an email on a familiar theme. It is from a colleague concerned about a junior researcher whose career decisions are “being twisted in an uncomfortable way” by the demands of New Zealand’s Performance-Based Research Fund (PBRF). “I have had many such messages over the years — which reflect the good and the bad of the PBRF,” says Boston.

Boston holds a personal chair in public policy at Victoria University of Wellington, and was one of the architects of the New Zealand system. He says that research paths can become conflicted by one of three scenarios: compulsion to publish articles in

high-impact international journals rather than working on a book with a domestic publisher; pressure to change research focus to better align with mainstream or more highly esteemed fields; or encouragement to accept a position as a non-PBRF-eligible teaching fellow and move away from active research. Each of these outcomes can be traced to the way in which the PBRF measures research excellence, and so its influence on the country’s research funding environment.

At its conception in 1999, explains Boston, the PBRF was an ambitious undertaking to measure research excellence and raise standards at institutions across New Zealand. It was intended to remedy years of neglect of the research sector through which the bulk of funds, dubbed research top-ups,

had been linked to postgraduate student numbers. The flaws of that arrangement had become evident in the mid-1990s when non-university higher education providers — such as polytechnics and institutes of technology — began offering postgraduate degrees. The research pot was suddenly being split between ever more institutions, many of which had limited research capacity. When Helen Clark’s Labour government came to power in 1999 with five former academics — including Clark, a lecturer in political studies — among its senior ranks, it vowed to strengthen the process of research funding and increase accountability. “The only option was some sort of performance-based regime,” says Boston.

The PBRF is based on the individual, making it unique among measures of national

research excellence. Every six years, it gauges and reports the standard of research of each of New Zealand's approximately 6,000 researchers in universities and colleges (so-called tertiary educational establishments). These rankings — A, B, C and R — are provided to the institutions; a researcher can apply to receive his or her own rating. The outcomes are then weighted by quality and subject area, in line with the resources required for different fields. The individual results are aggregated by institution and are the major determinant, alongside external income and research degree completions, of the distribution of research funding. The PBRF is now the largest single source of tertiary research funding in New Zealand, worth NZ\$262.5 million (US\$224.2 million) in 2013 (see page S52).

Across the Tasman Sea, the Australian Research Council (ARC) is gearing up for the third round of its own national measure of research quality, the Excellence in Research for Australia (ERA) evaluation, which takes place for every three years for the country's 41 universities. The forthcoming ERA2015 will categorize and evaluate the nation's entire higher education research output, comprising more than 400,000 publications. The ERA assesses work by discipline, and the initiative directly influences only a small portion of university research funding.

Results suggest that the overall quality of research has increased in both countries since the introduction of national assessments. More New Zealand researchers are achieving an A rating, and more Australian disciplines are classed at 'above world-standard'^{1,2}. But the PBRF and ERA prompt passionate reaction in their respective research communities. With increasing awareness of the need to assess the societal impact of research, merely weighing academic excellence makes less sense. There is concern that subject-focused assessment programmes don't adequately recognize the value of interdisciplinary research. And, as Boston's recent email correspondence implied, there are fears that the way excellence is measured — in particular the focus on high impact publications — may be hindering the careers of young researchers.

THE FOLLY OF YOUTH

PBRF and ERA both use metrics of quality and peer review to determine ratings. The same indicators are used in different forms in research assessment schemes around the world. However, for the PBRF, all nominated publications and other research outputs are rated by selected reviewers — a qualitative process that depends on individual judgements.

The ERA, by contrast, places more emphasis on citation analysis. "The number of people who are citing and making reference to work is a pretty good indicator of the significance and impact it has had in the academic community," says Aidan Byrne, ARC chief executive officer.

But these measurements of excellence are creating obstacles for young scientists, says Attila Brungs, deputy vice-chancellor for research at the University of Technology, Sydney (UTS). "Narrow metrics can drive some bizarre behaviours. People don't publish as much with PhD students because PhD students are often published in lower-ranked journals."

More broadly, the strengths of early-career researchers aren't readily demonstrated by reference to an objective publication review, a particular flaw of the individual-centred PBRF. Assessment encourages institutions to employ staff with established research records rather than emerging researchers who are doing excellent science but who are yet to amass publications. There is early evidence of this reluctance to engage young researchers, with one study showing a 14% drop in research staff aged 35 and younger between the first and second rounds of the PBRF³.

Indeed, a 2008 independent review of PBRF found that morale of otherwise high-achieving young researchers was being hurt by low ratings. The review, commissioned by the body that oversees PBRF, the Tertiary Education Commission (TEC), stated that "the assignment of a 'C' grade was seen by rising stars to undermine morale and to stigmatize their position". Boston says that when the PBRF scheme was designed there was no intention to reveal individuals' ratings. Not until after the system was established did Boston and his colleagues realize they were compelled to impart that information to researchers. "We simply failed to fully realize the implications of the Privacy Act and Official Information Act," he says. "If I had known we would end up with a regime in which individuals had their scores reported to them, and that other people could potentially know what they were, I would not have supported it."

The TEC has created a specific 'new and emerging researcher' category to counter disincentives to employ early career researchers when evaluation rounds loom. Researchers in this category can qualify for C(NE) rating and contribute to their institution's funding allocation. Their evidence portfolio assessment is weighted against their time as a researcher, with a minimum of two research outputs generally expected.

But many people, including Boston and Peter Gluckman, chief science advisor to the New Zealand prime minister, believe that the individual judgements inherent in the PBRF reviewing process continue to place undue pressure on emerging researchers to publish in high-impact journals. "I've seen several young

researchers quite compromised by this drive to produce the one paper that will get into *Nature*," says Gluckman, "when their career would have been much more developed had they focused on getting solid, excellent papers in the appropriate journals".

TOGETHER YET APART

Researchers undertaking interdisciplinary work are also feeling compromised. Campuses across New Zealand and Australia are bringing together researchers from multiple disciplines, from the hard sciences to the humanities, to look at societal problems in a holistic way. These fields include environmental sustainability and medical research and, in many cases, work is carried out under the auspices of a centre or an institute within a university.

Despite this big-picture approach, assessments such as the PBRF and the ERA continue to view research through a mono-disciplinary lens. The final report from the 2012 PBRF conceded that the 42 subject areas under which all research is assessed "do not accurately reflect the way research activity is organized and conducted". Despite this acknowledgment, there are no plans for a review, says Marny Dickson, chief policy analyst for tertiary education at the New Zealand Ministry of Education.

The story isn't much more encouraging in Australia. The Australian Council of the Learned Academies (ACOLA) — representing the four Australian academic societies: the Australian Academy of Science, the Academy of Social Sciences in Australia, the Australian Academy of the Humanities and the Australian Academy of Technological Sciences and Engineering — seeks to inform policy specifically related to multidisciplinary research. In a 2012 report, ACOLA found that the ERA "has difficulty in evaluating and reporting interdisciplinary research". And the situation is likely to be exacerbated as universities base their internal benchmarks around the ERA, which, like the PBRF, focuses researchers on higher impact journals — few of which are interdisciplinary.

For instance, the Centre for Cosmopolitan and Civil Societies at UTS, does a lot of applied research related to policy, and frequently produces work for the local, state and federal governments. When the centre's researchers publish, they have to do so in the journals of their individual expertise, whether marketing, business and economics, or social science and the humanities. The university is then rated separately for each of these fields, rather than for the centre's projects as a whole. "In an exercise like ERA, their work disappears, it doesn't exist," says Brungs.

Brungs says that the distortion doesn't yet affect the university's research priorities, because the funding linked to ERA is very small. In 2014 it was just AUS\$69 million, or 4% of the research block grants made by the Department of Education. But, an increase in

There are fears that the way excellence is measured may hinder the careers of young researchers.

INTERNALIZING TARGETS

Measurements used by the Excellence in Research for Australia system are finding their way into internal targets set by research institutions. Here is one example of 2011 targets from a leading Australian university.

Research outputs	Laboratory-based sciences: minimum (Min) and aspirational (Asp) targets							
	Lecturer		Senior lecturer		Associate professor		Professor	
	Min	Asp	Min	Asp	Min	Asp	Min	Asp
N° publications	1.5	4	3	6	4	8	7.5	15
Impact factor	4.5	16	9	24	12	32	22.5	60
Proportion in A/A* journals †	35%	55%	45%	65%	55%	80%	55%	80%
Research income (AUS\$)	\$5,000	\$40,000	\$30,000	\$125,000	\$150,000	\$500,000	\$500,000	\$1,000,000

† Ranking of journals as a proxy for quality (A/A* being the two highest) was abandoned by the ERA in 2011.

the proportion of ERA funding would make future collaborations harder to justify. “Universities are not allowing the drivers to distort their behaviour too much,” says Brungs. “But if we continue to go down that path it does have real danger for the interdisciplinary sector.”

PLAYING GAMES

If PBRF and ERA become more significant, recruitment policies at universities will inevitably be coloured by how a candidate might affect a pending evaluation. Some fear this will lead to a widespread gaming of both systems as institutions try to improve their scores.

Australia’s National Tertiary Education Union has found that ERA gaming already occurs, as individuals, departments and institutions strive for results needed to influence funding decisions that last for three years. In the case of the PBRF, the equivalent decision influences six years of funding, a magnitude which further incentivises manipulation. The problem is likely to persist, says Frank Larkins, former deputy vice-chancellor for research at the University of Melbourne. “Universities have a lot of smart people and they can learn pretty fast how to optimize their performance,” says Larkins.

The ERA peer-review panels are asked to look closely for idiosyncrasies in the research performance of institutions, and the ARC is now able to cross-reference dubious submissions against previous rounds. However, there is nothing to stop universities taking on researchers, and sometimes whole research departments, in order to boost output prior to an ERA round. The Australian newspaper described the “churning” of researchers this year in the lead-up to the 31 March 2014 census deadline. Any staff hired after this date are not eligible for assessment in ERA 2015. But, during the preceding Australian summer, research groups and even whole departments were poached by the Australian Catholic University, Central Queensland University and Charles Sturt University, among others.

ARC leader, Byrne, doesn’t endorse such activities, but says that calculated reallocation of resources for the purposes of ERA ranking is not necessarily a bad thing. “We don’t want to stop institutions from making

strategic decisions about what research they wish to pursue.” The importance placed on the ERA rankings by Australian university management has been evident, not only in the tendency for researcher churning, but also in the tailoring of internal research benchmarks to better meet the terms of the ARC system (see ‘Internalizing targets’). Consequently, universities have set departmental and school-wide targets regarding quantity and quality of publications.

Despite this, Byrne does not accept that ERA is forcing institutional change. “It does get used by institutions in various ways, but we are providing an evaluation against the best possible standards we can come up with”, he says.

Perhaps the most egregious example of an attempt to game the system occurred in New Zealand in 2006. One leading university reclassified dozens of staff members, notably those who were PBRF-eligible but performed little active research. By reclassifying inactive researchers away from subjects such as economics and biology to fields such as philosophy and religious studies, the university would improve its standing in the former fields. The surge in the number of New Zealand philosophers piqued the curiosity of PBRF reviewers who eventually reversed the classifications.

MEASURING A MOVING TARGET

The objectives and the structure of ERA and PBRF have changed little, but the status quo may be threatened by demand for the explicit inclusion of research impact as a quality indicator within the assessment exercises (see page S81). At UTS, Brungs says that more focus on impact might bring much-needed formal recognition of interdisciplinary work within the system. For example, work that has valuable outputs concerned more with policy than scholarship. “Publishing in *Nature* is one way of demonstrating excellence in research,” he says. “Changing the way that a nation drinks water is another way.”

The ARC is considering the inclusion of impact measurements, but Byrne says the organization does not want to just graft these on to the existing system and does not have the resources to develop an independent

measurement of impact. The Australian government’s current aversion to any increase in red tape does not help.

In New Zealand, the 2008 review of PBRF cautioned against diluting its focus on excellence by aligning it with government innovation policy. However, a re-evaluation has seen a number of alterations, including an increase in the significance of investment from industry in determining overall funding awards, coupled with a moderate reduction on the emphasis of the research quality assessment. According to the Ministry of Education, these changes reflect the fact that external research income is a “strong proxy indicator” for the transfer of knowledge between academia and industry and the change will encourage “research of relevance to end-users.” It is a tangible shift towards reward for research impact.

But Boston is not satisfied that, even with such changes, assessments like the PBRF and ERA will continue to be relevant. He refers to Goodhart’s law, which states that once a measure becomes a target it ceases to be a good measure. “I don’t see the logic of running the same assessment process every six years ad infinitum, with only minor tweaks,” he says. “It sets up a particular set of incentives and a particular kind of process within institutions, some of which is undesirable.”

To keep improving research excellence, Boston says, the government needs to increase funding and other resources, or make bigger changes to the assessments, for instance by introducing new criteria. Continual re-allocation of finite resources can only do so much. “I don’t know how you’re going to squeeze more drips out of the orange.” ■

Annabel McGilvray is a freelance science and medical writer based in Sydney.

1. *Performance-Based Research Fund: Evaluating Research Excellence — the 2012 Assessment, Final Report* (Tertiary Education Commission, 2013); available at go.nature.com/gprpet.
2. *Excellence in Research for Australia 2012, National Report* (Australian Research Council, Commonwealth of Australia, 2012); available at go.nature.com/q8sxiw.
3. Çınlar, N. & Dowse, J. *Staffing and Performance Trends in Research Subject Areas* (Tertiary Education Commission, 2008).