

Professor Stuart Macintyre

**‘ERA, Overview and Implications: The Academic Perspective’
NAF Workshop on Excellence in Research Evaluation: Impacts on Scholarship,
Research Funding and Publications
Stuart Macintyre,
President Academy of the Social Sciences**

In the closing years of the twentieth century, most advanced industrial countries adopted similar procedures to evaluate university research performance. They did so as part of a larger endeavour to maximise the contribution of higher education to their innovation system.

Innovation has been described as novelty with an economic value, and the novelty is found in products and services, the processes whereby they are produced and the organisational arrangements surrounding the productive process.

A national innovation *system* is taken to be an ensemble of the country’s public and private sector institutions and activities. Universities have a particular role in such systems. They provide the education that increases the capacity for innovation; they train the knowledge workers who lead innovation; and they contribute to knowledge, especially in basic research, which is a vital component of an innovation system.

Australia comes late to the assessment of university research performance. It was not until the late 1980s that we separated the funding of teaching and research, and created an ARC with responsibility for providing research funds on a competitive basis. The ARC began with a significant emphasis on evaluation of its programs, and its early efforts included stocktakes and disciplinary reviews. But these activities dwindled in the reorganisation of the ARC in the late 1990s, and the agency lost its capacity to undertake them.

It has now been charged with conducting a national assessment and acquired staff from the Department to assist it; my own feeling is that it needs an additional capacity for finer-grained evaluation.

Following the establishment of the ARC, the Commonwealth government created a set of measures for the allocation of funds for research performance at the institutional level, but these were little more than a reflex of grant income. That income was largely determined by the operation of just two agencies, the ARC and the NHMRC, overlaid by the formulation of national priorities and special initiatives at the behest of PMSEIC and the Chief Scientist. By international standards, this was a very narrow basis of assessment

Belatedly, the last government decided it would conduct a more systematic assessment, the RQF, and universities devoted substantial resources to preparing for it. The present government scrapped the RQF in favour of the present ERA, which was conceived as a ‘light-touch’ alternative.

What is the reason for such assessments? They arise in part as a form of management designed to satisfy expectations of accountability and efficiency. Beyond that, they are tools for government to align its support for university research to performance. In a mass system of higher education, the public provision of research funding can never meet demand, so along with competitive processes to award funds go processes to measure quality, both assisting government to concentrate a scarce resource.

Research assessment has thus been adopted by the majority of OECD countries as an instrument of their national innovation systems — with one conspicuous exception. The United States, whose high-tech industries are widely taken as a model of knowledge transfer, has a complex variety of funding mechanisms for university research, and various ways of assessing performance — but has not undertaken a national assessment. By its nature, it prefers plurality to uniformity.

Granted, the United States has the luxury of possessing the largest economy and the richest system of higher education. Other countries lack its capacity and feel a need for closer control. But in seeking to emulate America's success, its imitators are inattentive to its methods.

The declared purposes of ERA are

1. To provide an evaluation framework that gives government, industry, business and the wider community confidence in the excellence of research conducted in universities
2. To provide a national stocktake that will identify areas of excellence, emerging research areas and opportunities for development
3. To allow for national and international comparison.

All of these are purposes that are difficult to gainsay. As the Minister seeks greater budgetary provision for research, he needs to be able to present evidence to ministerial colleagues and the public that the research activity of Australian universities deserves it. He will be assisted if he can identify areas of excellence and points of growth. Without comparison, the evaluation will have little credibility. I'm not sure that the design of ERA will allow these purposes to be met, and we shall no doubt discuss the adequacy of its methodology.

The initial announcement of ERA stated that it was not to determine funding. This was a reassurance that Australia would not follow the example of the British RAE or the New Zealand PBRF, where outcomes were tied directly to funding arrangements. Yet it was apparent from the outset that the outcomes of ERA would have significant consequences.

One early indication was that the ERA results would inform the minister's negotiations with universities on compacts covering the range and extent of their research activity, with the idea of hubs and spokes arrangements whereby research in particular disciplines would be linked and concentrated. Another was the suggestion that ERA outcomes would

affect higher degree research, so that where research performance in particular disciplines was below a threshold level they would no longer participate in the Research Training Scheme.

More recently, the government has released a discussion paper on Sustainable Research Excellence, linked to the provision in the 2009 federal budget for making good the shortfall in the indirect costs of university research. The issues paper proposes that most of the funding be contingent on the university meeting performance thresholds and participating in ERA, with banding of performance in discipline clusters the principal determinant of allocations.

These consequences have clear implications for the ERA methodology. It is one thing to produce a dashboard of indicators for research performance in a particular field of research, another to arrive at rankings that will determine the future of that field at an institution.

If the La Trobe philosophers learn that they lag behind their colleagues at Monash in various indicators of research excellence, you can be sure they will exercise all their considerable ingenuity to refute the slur. But if they are deemed to be below the threshold for, say, research training, while their colleagues at Monash are above it, then they will demand evidence that the judgement is justified. I'm not sure that the present ERA methodology allows that demand to be met.

ERA is to use indicators of quality (ranked outlets, citation analysis, peer reviewed income), volume of activity and indicators of application, with esteem factors withdrawn from the trial evaluations and still under development for the 2010 exercise. In many of the Humanities and Creative Arts and now Social, Behavioural and Economic fields of research, where the patterns of research publication do not allow for citation analysis to yield reliable results, there will be a component of peer review.

My own view is that a dual system of evaluation is needed, especially for the first round of ERA. Here I am influenced by the work of Linda Butler (as well as the evidence provided by David Sweeney yesterday) that shows significance variation in the results provided by peer assessment and bibliometric evaluation.

Much of the argument during the preparation of ERA has centred on the ranking of outlets, its validity and consequences; hence the special attention at this symposium on the implications for scholarly publication. Many scientific disciplines have operated for some time with journal rankings based on citation, but the rankings for ERA use a different method, which might variously be described, according to your views, as expert evaluation or subjective judgement.

In preparation for ERA advice was gathered on the assignment of journals in particular fields of research to four categories, from A* to C: A to D would be a less euphemistic taxonomy.

The initial rankings attracted criticism. In a number of the humanities and social sciences, fears of a cultural cringe led to what A.A. Phillips described as the obverse, cultural strut: many worthy but far from outstanding Australian journals were given higher rankings than strong international ones, and I am sure I not the only person here to have received communications from the editors of those international journals, mixing lamentation and indignation.

The ARC has devoted considerable effort to revise the lists for the HCA trial evaluation, and no doubt will make further revision before the full exercise — it is to be hoped that universities understand the trial evaluation is indeed a trial, and not use the results for premature judgement.

The problems with journal ranking have been widely canvassed. They include the risk of underestimating journals in emerging fields, the danger of neglecting specialist journals and the problem of bias towards English-language journals. The rankings for humanities journals produced by the European Science Foundation came in for particular criticism and failed.

There is also the danger that a measure becomes a signal, driving researchers to submit their articles to A* journals regardless of their appropriateness, and draining support from other journals.

The practice of deans and heads directing their staff to publish only in top-ranked journals is already apparent, and has the perverse effect of weakening the Australian presence in the literature of their field.

The journal rankings are only one of the indicators that bear on the avowed objective of ERA to compare the quality of Australian research with international research. Another is the use of the Australian and New Zealand Research Classifications as the unit of analysis. In both cases the decision to use national measures is surely correct but only emphasises the contextual character of evaluation.

The point has been often made that international statistics such as those produced by the OECD ignore different national practices in defining, arranging and classifying research activity. The same holds for any use of ERA to identify Australia's research excellence: it will require closer attention to disciplinary configurations than the quantitative evidence generated by these classifications.

Comparisons of excellence across fields of research will also be difficult, for ERA recognises disciplinary differences, and has accepted that no single indicator can be applied across them. Its willingness to develop appropriate measures for different clusters is commendable, but limits comparability. It is to be hoped that we avoid the outcome of the first NZBRF, which relied heavily on a narrow base of peer review and allowed a newspaper there to report the outcomes under the headline 'Our Philosophers Are Tops'.

Like other research assessment exercises, ERA also rests on assumptions about concentration of research activity. It sets minimum levels of publication in a field of research in a four-digit Australian and New Zealand Research Classification as a condition for evaluation, and its likely consequences are designed to recognise, reward and consolidate present concentrations of effort.

Yet the advantages of scale remain more an article of faith than a demonstrated effect of research policy, and across disciplines the minimum requirements of equipment and personnel vary widely. It would be a pity if small groups of research excellence were to be disadvantaged, indeed it would increase the isomorphism of the Unified National System.

The same holds for research that is less than excellent. Many universities have departments and schools that undertake good but not outstanding research. Such work often has links to its locality, is frequently applied in character, sustains modest but viable honours and higher degree research programs, but is unlikely to be rated highly on ERA's indicators. Many of these research activities are undertaken in regional universities, and are part of a fragile ecology of teaching and research.

We should recall also that our system of dual funding differs from the UK. General University Funding in Australia makes less provision for research. Our geography and the more limited mobility of students make it harder to rely on specialisation.

To say this is to remind ourselves that Australian higher education is still wrestling with the contradictions of a Unified National System that was created twenty years ago, one that has no counterpart in any other country in that it consists entirely of universities that seek to be comprehensive, with government policies that have repeatedly proclaimed diversity and yet operated with arrangements that encouraged conformity — and furthermore, provided inadequate levels of public support. I'm not at all sure that ERA is the appropriate mechanism to break this deadlock.

I welcome the replacement of the present system of measuring and rewarding research through the Institutional Grants Scheme, with its rigid input measures and limited output measures, especially the publications formula that had such pernicious effects. I welcome the attempt to devise an alternative, the placement of it with the ARC and the spirit with which it has been developed. But it remains a work in progress and an unproven tool of measurement with implications that invite further consideration before it is used as a basis of policy.