The Allen Consulting Group

Establishing an Australian Research Quality Framework

Summary of discussion from the Australian Academy of Science Workshop

March 2005

Report to The Australian Government Department of Education, Science and Training

The Allen Consulting Group

The Allen Consulting Group Pty Ltd

ACN 007 061 930

Melbourne

4th Floor, 128 Exhibition St

Melbourne VIC 3000

Telephone: (61-3) 9654 3800

Facsimile: (61-3) 9654 6363

Sydney

3rd Floor, Fairfax House, 19 Pitt St

Sydney NSW 2000

Telephone: (61-2) 9247 2466

Facsimile: (61-2) 9247 2455

Canberra

Level 12, 15 London Circuit

Canberra ACT 2600

GPO Box 418, Canberra ACT 2601

Telephone: (61-2) 6230 0185

Facsimile: (61-2) 6230 0149

Perth

Level 21, 44 St George's Tce

Perth WA 6000

Telephone: (61-8) 9221 9911

Facsimile: (61-8) 9221 9922

Brisbane

Level 11, 77 Eagle St

Brisbane QLD 4000

PO Box 7034, Riverside Centre, Brisbane QLD 4001

Telephone: (61-7) 3221 7266

Facsimile: (61-7) 3221 7255

Online

Email: info@allenconsult.com.au

Website: www.allenconsult.com.au

Chapter 1

Australian Academy of Science Workshop

1.1 Workshop details

The Australian Academy of Science (AAS) workshop was held on Friday 11th February, from 10am until 2.30pm, at Ian Potter House, Gordon Street, Canberra.

AAS participants at the workshop were:

- Professor Bob Frater – Vice-President for Innovation, ResMed
- Professor Jenny Graves – Director, ARC Centre for Kangaroo Genomics
- Professor Philip Kuchel – McCaughey Professor of Biochemistry, University of Sydney
- Professor Bruce McKellar – Professor of Theoretical Physics, University of Melbourne
- Dr John Passioura – Honorary Research Fellow, CSIRO Plant Industry
- Professor Sue Serjeantson – Executive Secretary, Australian Academy of Science
- Professor John White – Professor of Physical and Theoretical Chemistry, Research School of Chemistry, ANU
- Professor Bob Graham – Executive Director, Victor Chang Cardiac Research Institute

Ms Lyndal Groom and Mr Steve Kozel from DEST attended in an observer capacity.

Dr John Bell and Mr Duncan Buckeridge from The Allen Consulting Group facilitated the workshop.

1.2 Workshop agenda items

The workshop was structured into two sessions.

Following a brief introductory statement by Mr Steve Kozel (DEST), the first session, dealing with issues surrounding what are appropriate high level attributes for a Research Quality Framework (RQF), addressed the following questions:

- What are the appropriate purposes for an RQF?
- How should the quality of research be defined?
- What is the appropriate unit of assessment within an RQF?
- What types of activity should be assessed within an RQF?
- What institutions should be included within an RQF assessment process?
- How frequently should assessment occur?

The second session explored the following implementation issues for an RQF:

- What are existing quantitative and qualitative metrics for quality and impact for research in the technological sciences and engineering?
- Are additional quality measures needed?
- If a peer review panel system was to be introduced, how many different panels would be needed to cover the fields of research encompassed by AAS?
- How can future/potential/emerging excellence as well as past excellence be recognised?
- Can the National Research Priorities be applied within an RQF?
- What sort of transition arrangements would be needed with introduction of a RQF?
- What are the outcomes from an RQF process that should be definitely avoided?

The workshop then concluded with some brief remarks and a thankyou by Mr Steve Kozel (DEST).

1.3 Proceedings of workshop

Session One

What are the appropriate purposes for an RQF?

At the outset of the discussion of appropriate purposes for an RQF, it was made clear that participants believed that the introduction of an RQF would be ineffective if it was not linked in some way to funding allocation. It was suggested that the current mechanisms for allocation of block funding to institutions were seriously flawed and needed to be changed – the Institutional Grants Scheme in particular was consistently criticised as not promoting quality.

However, it was also noted that a focus on funding allocation may limit the extent to which an RQF can genuinely focus on improving quality since once funding is involved that is all anybody will really focus on.

In addition to serving a funding allocation purpose, it was also generally agreed that an RQF should also be designed so as to encourage better performance within the research system.

Another point raised was that in addition to explicitly identifying where quality research is being generated, a purpose for an RQF would be to explicitly identify where quality research is not being generated.

How should the quality of research be defined?

One challenge in defining quality was seen to be that for different purposes or types of research quality may mean different things. Successfully embracing the full spectrum of research, on both a sectoral (different types of institutions with varied missions) continuum and a research type (basic to applied research) continuum was seen to be an important feature for an RQF. However, there was general agreement that the concept of excellence, while perhaps being judged differently in different fields/types of research, was central to the definition of the quality of research.

The proposition was put to stakeholders that perhaps one way to resolve this challenge would be to assess research quality in terms of whether a particular piece of research had achieved its particular stated goals. However, participants raised a an important difficulty associated with the use of such an approach. Research outcomes were seen as often hard to anticipate, and while stated research objectives may not be achieved, other quite different – yet more 'valuable' (not defined monetarily) – outcomes may have been achieved. If research quality was to be defined in terms of success against pre-stated goals, it was agreed that much excellent research would be discounted.

Inclusion of the concept of relevance (alongside excellence) in the definition of quality was seen as challenging to deal with in practice. It was suggested that while excellence and impact of research may be linked, the time scales on which they can be observed and assessed will often be very different. The final impact or relevance of excellent research may not be apparent for many years after the research is done. It was also pointed out that the citation half-life of publications may differ greatly across different fields of research – possibly indicating that in some areas it takes longer for impact to become apparent than in others.

Another approach to defining quality that was raised was the prospect of explicitly assessing the quality of research leaders (methodological difficulties notwithstanding). Accompanying this view was the suggestion that the best way to allocate funding is to 'put money into the good leaders'. It was noted that a high focus on identifying (and funding) high quality leaders would likely see an increase in the mobility of such leaders between institutions. This may or may not be problematic.

What is the appropriate unit of assessment within an RQF?

Competitive grant schemes (such as those of the ARC and NHMRC) were seen to already focus on the evaluation of individuals. A suggestion raised was that an RQF, if it is to be used to inform allocation of block funding, should not focus on individuals as the unit of assessment.

Using the department/faculty as the unit of assessment within an RQF appeared to have the highest level of support amongst participants. An alternative approach with some support was that institutions could be asked as a whole to put forward the best examples of research output from within their institution within particular research field groupings.

What types of activity should be assessed within an RQF?

Most discussion on this question surrounded whether research training should be included alongside research outputs within an RQF. One view that received considerable support was that training should be included within an RQF to ensure that the standard of research doctorates is protected. There was the suggestion that, for example, the PhD standard is in danger of being eroded – in part as a consequence of current funding formulas giving high incentives for institutions to generate high numbers of PhD completions.

There was then some discussion of how research training quality could be assessed. Effectively doing this appeared to be a difficult task. Process based approaches were seen as likely to lead to quite bland results, while outcomes measures such as student destinations may not capture the value-added being provided by institutions – student destination may be driven as much by the quality of students admitted to an institution as by the quality of training they receive at an institution. It was also noted that some students, due to personal circumstances rather than ability, may not be able to play the global job market. Also, it would need to be recognised that there are a range of valued student destinations – academia, government, business – and the mix may differ by discipline area or type of research focus (on the basic to applied continuum) of an institution or faculty.

Another possibility raised was the potential to include commercialisation outcomes as a category of output to be assessed within an RQF. However, following a brief discussion of this, a general consensus emerged that commercialisation outcomes are not the same as research quality and the way that they would need to be assessed would also require more of a focus on processes than on assessing actual outputs. A common view raised was that while, for strategic priority reasons, there may be a desire to give more funding to commercial activities, this does not reflect the quality of commercially focused research vis-à-vis the quality of non-commercially focused research. It was therefore generally agreed that commercialisation outcomes should not be considered within an RQF.

What institutions should be included within an RQF assessment process?

There was general agreement that it is desirable to include as many public research institutions within an RQF as possible. However, an important proviso raised was that this was true provided there was no attempt to 'claw back' funding from all participating institutions across the research system to form a common funding pool to be allocated via an RQF assessment.

It was also suggested that the inclusion of a diverse range of institutions within an RQF would have several important implications for the operation of an RQF, namely:

- that it would preclude linking outcomes to funding via a simple common formula; and
- that it would preclude pre-weighting of categories of outcomes into a common formula for calculating a final assessment score.

An approach that was generally endorsed was to include all players within the research system in an RQF but then use diverse ways of judging performance and keep funding pools separate for different categories of institutions (as is currently the case).

How frequently should assessment occur?

It was suggested that if an RQF was not to be linked to funding in a formulaic sense, there may be a need to only conduct it once – as a test of how performance now looks so that the appropriateness of current funding allocations can be assessed, and, if necessary, re-thought.

If an RQF is to be linked to funding allocations in a formal way, it was agreed that an RQF would need to be repeated at a given interval. While a number of factors, in particular the scale of administrative burden associated with an RQF, will effect what is the most appropriate frequency for the assessment rounds, the most common figure floated was a frequency of once every five years.

Session Two

What are existing quantitative and qualitative metrics for quality and impact for research in the sciences?

A large number of quantitative and qualitative indicators were recommended so as to accurately recognise, encourage and reward excellence within a diversified research system. Particular existing available quantitative bibliometric measures raised as potentially useful measures, included:

- publication counts;
- impact factor adjusted publication counts;
- citation counts; and
- impact factor adjusted citation counts.

A number of well established methodological challenges associated with any such measures were then raised – opportunity for 'gaming' of the system, self citation, higher citation rates in some disciplines than others, high citation rates for particularly bad papers, disputes over impact factors, and so on.

The methodology of using some type of survey of the research community to establish leaders in particular disciplines was not supported by participants as it was seen as being vulnerable to gaming and for being unlikely to pick up emerging excellence. The importance of leaders and the notion that funds should be directed to such identified leaders was nonetheless agreed.

Given the difficulties associated with any particular metric for quality, it was generally agreed that it would be necessary to allow for a bundle of measures to be used and that peers/experts would then need to 'sift through the bunches of evidence and pick out the gems'.

Are additional quality measures needed?

It was not felt that new quality measures would be needed, rather, it would simply be necessary to allow a wide range of existing measures to be included in the portfolios of evidence that would be put forward for expert review.

The prospect of such portfolios of evidence being subject to a 'probity audit' – i.e. that claims made in portfolios would be subject to random verification checks – was rejected. Such a system was seen as likely to be damaging to morale, hard to carry out in practice, and unnecessary – as those groups submitting portfolios of evidence to expert or peer panels would know that they could not get away with misrepresenting achievement levels.

If a peer review panel system was to be introduced, how many different panels would be needed to cover the fields of research encompassed by AAS?

It was generally agreed that an RQF would need to involve some system of expert review of research outputs – both to accurately assess quality and to have credibility within the research community. It was felt that exclusive reliance on proxy measures for quality, such as bibliometrics, would, while perhaps acceptable in some disciplines, be generally inadequate and lead to dubious assessment outcomes.

A key question discussed would be whether it would be necessary for research outputs to be assessed by a panel of peers with specific expertise in the field of research that is being assessed or whether it would be satisfactory for a more general panel of experts – drawn from a number of specific disciplines – to make assessments of research quality. An analogy was drawn to the way that fellows are elected to learned societies, where experts, but not necessarily discipline peers, make the final judgement. However, it was noted that such expert selection panels would generally seek input on applications from discipline peers of the candidate.

A problem raised by one participant in relation to reliance purely on peer panels was that sometimes disciplines become so focused on how things are done in that discipline that people in the discipline may loose the ability to 'see the forest through the trees'. The use of expert panels, where a broader set of perspectives may be brought to bear on the assessment of research outcomes, was seen as a way to avoid a tendency to disciplinary insularity.

Conversely, there was the view expressed that without discipline specific expertise it may be very difficult to assess the quality of research outputs.

An overall consensus appeared to emerge towards the end of discussion on this issue that some type of blended panel system would be needed. Under this approach, panels would need to have enough discipline expertise to assess research outputs, but enough breadth of perspective to ensure consistency in assessment across specific discipline areas and to allow for multi-disciplinary research outputs to be properly assessed. Getting the mix right was seen as the challenge. Also, the more discipline specific expertise that was needed to make assessments, the greater the number of panels would be needed.

It was also noted that on a practical level, the number of panels would be driven to a fair degree by what is an acceptable workload for any panel to undertake. There would need to be enough panels to ensure that the workload for panel members is not excessive. The potential was raised to convene panels under three headline clusters: the humanities (taken to include the social sciences); medical and biological; and the technical sciences. Under each cluster there would then be as many panels as is necessary to handle the workload.

In terms of actual panel numbers, another view was that there should be as many panels as can be afforded.

In terms of the composition of panels, the inclusion of some overseas members was seen as desirable for two reasons: it would bring an international perspective to assessments and it would also serve an ambassadorial function as those exposed to the high quality of Australian research would take that knowledge home with them at the conclusion of the process and 'spread the word'. The inclusion of some people from outside the academic community on panels was also seen as being potentially beneficial – particularly from the perspective of increasing the credibility of outcomes with Treasury.

Another practical issue raised was that, in order to be useful as a quality improvement tool, panel reports should ideally be 'warts and all'. However, there may be a risk of litigation by people who disagree with panels' conclusions, so panels may not want to release full and frank reports. Some protection for panels to ensure that reports are not self-censored may be needed for the system to work effectively.

How can future/potential/emerging excellence as well as past excellence be recognised?

The inclusion of 'early career researcher performance' amongst the menu of self-assessment items to be considered, was seen as a way of ensuring that emerging excellence is captured within an RQF process. Ensuring that the nurturing of the next generation of researchers was identified and encouraged was seen as an important purpose for an RQF.

Can the National Research Priorities be applied within an RQF?

The common view was that whether research bears on a national research priority is not relevant when assessing the quality of that research. The common view was that the National Research Priorities should not be incorporated within an RQF.

What sort of transition arrangements would be needed with introduction of an RQF?

Participants agreed that transition arrangements (in terms of stakeholder support) would be difficult if an RQF was to be used to simply reallocate existing funding. However, the transition could be expected to be smooth if additional funding was provided for an RQF.

One option put forward was to conduct a test run with no funding attached, explore what the results tell you, and then run (within a year or two) a second round with funding attached.

A contrary view, that received a fair degree of support from participants, was that for the exercise to be taken seriously it should have significant funding attached to the first round. However, funding changes would have to be incrementally introduced over several years to allow for the system to adjust in an orderly way. Accompanying the suggestion that significant funding should be attached to the first run was the proposition that this should be linked to a significant increase in overall funding being made available. If funding overall was increased significantly, putting half of all block grant allocations up for reallocation via an RQF would be appropriate. The suggestion was that unless such a significant of funding was involved the exercise would not be worth doing.

In relation to funding pools across different elements of the research system, it was felt that funding pools should not be combined, but rather, kept separate as is currently the case. Different parts of the research system were seen to have quite different missions and it was felt that it was a strategic decision as to how money is spread across different missions rather than something that should be determined by a quality assessment system.

There was the suggestion that it would be sensible to model the funding implications of a range of 'mock assessment measures' prior to implementation of any new funding system based on an RQF assessment.

What are the outcomes from an RQF process that should be definitely avoided?

While time constraints prevented any lengthy discussion specifically focused on what an RQF process should avoid, some points raised were:

- that the system should not be designed in such a way as to produce assessment outcomes that at first glance present the research system's performance in an artificially negative way. The experience in New Zealand, where as a result of the scoring system adopted no institution scored more than 4 out of 7 and the average score was less than 2, was highlighted as an example of the type of basic outcomes reporting design error that can lead to very damaging consequences. (It was noted that only a very small number of institutions globally would have scored above 5 or 6 using the New Zealand scoring system);
- the system should not lead to safe research at the expense of risky research; and
- the system should not encourage a focus on short-term outcomes at the expense of long-term outcomes.

At the conclusion of the session a more comprehensive written list, prepared by the Academy, of outcomes that an RQF system should avoid, was tabled. This list is attached as Attachment 1 to this summary of workshop discussions.

1.4 Summary of overall themes from AAS workshop

Major themes to emerge from the AAS workshop were that there was general acceptance of the proposition that the introduction of an RQF could play a useful role in:

- better allocating discretionary institutional funding within the research system; and
- helping boost the quality of publicly funded research in Australia.

The tone of the workshop was one of open and constructive engagement with the process, with a view to achieving the best possible outcomes in relation to the structure and conduct on an Australian RQF.

There was a consensus of opinion that an RQF should be linked to funding outcomes and that without such a link the exercise would be a waste of time.

There was a general view that an RQF should focus primarily on the quality of research output rather than process issues relating to research management. It was felt that an RQF should be primarily focused on identification of where quality outputs were and were not being delivered within the research system.

A model of faculty/departmental-level assessment involving self-assessment against a menu of performance areas (accompanied by guidance on appropriate metrics and review by blended expert/peer panels) had broad support as a potentially effective and workable model for an RQF.

Participants agreed that the more additional funding offered alongside the introduction of an RQF, the less significant any transition issues would be, while, conversely, if an RQF was to be used to simply reallocate existing funding, transition issues would be significant.

It was also stressed that continued stakeholder engagement and input into the design of an RQF would be crucial for its acceptance within the research community.

Attachment 1

List of things for an RQF to avoid provided by the Australian Academy of Science

At the conclusion of the workshop, a list of things that an RQF should avoid was provided by the Academy. Separate lists of things to avoid when assessing research quality and research training were provided. These lists are presented below.

When assessing research quality, things suggested for an RQF to avoid were:

- encouraging 'safe' research. Early and mid-career researchers especially were seen to need to take risks to break new ground. Researchers who contrive to extend a technique or approach which guarantees publication are not seen to be extending knowledge;
- introduction of rewards that foster short-term thinking;
- introduction of policies that reduce innovation and dynamism, and independent and original thought;
- international 'cringe'. An effective assessment of quality would be expected to show that Australian research is in many instances world-leading. Showing this was seen as a way of giving Australians assurance of the quality and value of Australian research;
- introduction of assessment procedures that are expensive, time-consuming, unnecessary or opaque;
- too many inputs and excessive focus on inputs;
- multiple counting of the same indicator;
- the treatment of every discipline, institution or sector in the same way;
- homogenising the Australian institutional mix in the interests of simplicity;
- a lack of focus on aims;
- conflict with institutional governance arrangements;
- game playing and abuses of the system;
- the creation of a pretext for major and abrupt changes to the funding of an institution;
- too much retrospectivity;
- overlooking or discounting the value of niche or regional universities and research institutions;
- excluding small or unusual research activities; and
- encouraging complacency in large and well funded institutions.

When assessing research training quality, things suggested for an RQF to avoid were:

- introduction of policies that reduce innovation and dynamism, and independent and original thought;
- introduction of assessment procedures that are expensive, time-consuming, unnecessary or opaque;
- too many inputs and excessive focus on inputs;
- introduction of rewards that foster short-term thinking;
- the treatment of every discipline, institution or sector in the same way;
- a lack of focus on aims;
- conflict with institutional governance arrangements;
- game playing; and
- encouragement of training for training's sake.