RESEARCH QUALITY FRAMEWORK DISCIPLINE WORKSHOPS  
FEBRUARY/MARCH 2007

Combined Workshop Summary

Following the release of the Development Advisory Group’s Recommended RQF in November 2006, the Department of Education, Science and Training (DEST) called for nominations to participate in Discipline Workshops. The primary purpose of the Workshops was to gather input from nominated sector representatives to assist the development of Research Quality Framework Specifications.

Of the nine hundred nominations received, just over one hundred representatives, drawn from a range of institutions, were invited to participate. Workshop Participants were chosen on the basis of ensuring adequate representation of the broad disciplines within each of the 13 Panel areas.

The schedule of seven workshops was comprised by grouping some Panels together according to the similarity of research outputs and shared issues around the measurement of research quality and impact. Panels 11 and 13 were dealt with individually in recognition of the distinctive nature of the research outputs in these areas.

The Workshops were facilitated by Emeritus Professor Fred Smith, who is currently assisting DEST in drafting RQF Specifications, and Ms Linda Butler and Dr Claire Donovan, who chaired, respectively, the RQF Development Advisory Group Quality Metrics and Impact Working Groups in 2006.

Facilitators invited participants to provide input on elements of the Framework identified as requiring discipline-specific detail, including: appropriate research outputs for quality and impact; research group size; quality metrics and what constitutes evidence of impact. Participants were also asked to comment on any other areas of particular concern for their discipline.

Eligibility

Issues around the criteria for eligibility for inclusion in the RQF were raised at all workshops with many supporting the inclusion of Level A Researchers. Representatives of the Humanities, Arts and Social Science (HASS) disciplines expressed a particular concern; they argued that disciplines in this area were likely to have a greater proportion of researchers employed at Level A.

Participants representing the clinical sciences and clinical physiology raised concern over the 0.4 FTE employment criteria. Some participants argued for an exception to be made for adjunct/conjoint appointments in university medical schools on the grounds that while conjoint and adjunct research were not employed by the university, they made a significant contribution to the clinical research undertaken by the university.

This was noted, but it was pointed out that the Framework did allow for the inclusion of research undertaken jointly between an eligible researcher and a conjoint/adjunct researcher. Such research could be included in the Context Statement and/or on a research output jointly authored with an eligible researcher.

Representatives from other panel areas raised the issue of technical staff and library professionals, who made a significant contribution to the research of a group. It was argued that the exclusion of these groups could serve to discourage them from being research active.
Research Groups

All workshops raised the issue of Research Group size. Participants were asked for panel-specific input on the justification for submitting a Research Group comprising fewer than five researchers and in this context participants sought clarification on whether there would be any disadvantage if a Research Group failed to demonstrate sufficient coherence.

Participants from the HASS sector voiced concern over the perception of a ‘Group’ as a cohesive research unit. They said that while the use of the term in this way may be appropriate for many laboratory-based sciences, much of the research undertaken in their area was done by very small groups and/or individuals. Representatives of Mathematics also argued that a lot of the highest quality research in Mathematics happens at an individual level.

The general view was that the majority of claims made for a low group number would be made on the grounds of disciplinary practice. Smaller and newer institutions would also be more likely to put forward Research Groups with fewer than five researchers.

Participants agreed that the issue of whether Groups would be disadvantaged by submitting a small group needed to be examined more closely.

It was noted that the Framework accommodated exceptions to the Group size criterion, but participants strongly urged that exemptions be considered by Assessment Panels well in advance of the submission deadline.

Acceptable Research Outputs

The issue of what constitutes a research output for the purposes of the RQF was discussed at all workshops, with views of participants varying greatly between, as well as within, Panel areas.

There was general agreement from participants that a research output could only be considered for the Quality assessment if the date at which it was finalised had occurred within the 6-year assessment period (1 January 2001 – 31 December 2006).

While participants acknowledged that some parameters around period of publication were necessary, concerns were raised regarding the long lead times in some disciplines between finalisation of the output and its publication. They also raised concerns about performances that moved from a relatively obscure venue to a national and/or international venue. If venues were to be rated, obviously researchers would prefer to report the national/international performance, even though the community-based performance was essentially the same work.

Cross Institutional and Multi-disciplinary Research

The issue of how cross-institutional Research Groups would be attributed was raised by a number of disciplines. It was argued that long-standing cross-institutional collaboration needed to be encouraged. Concerns were raised over how research of this nature would be assessed, either in a cross institutional Evidence Portfolio or in multiple cross-referenced Evidence Portfolios. There was agreement that the two options should be trialled.
Participants also raised concern over the assessment of multi-disciplinary research and the possible disadvantage that might arise from Panels other than the Home Panel contributing to a Research Group’s assessment. There was general agreement by participants in all workshops that Research Groups should have the discretion to nominate the Home Panel. They argued that allowing Groups the discretion to nominate which Panel judged their work would help address any short-comings with the Research Fields, Courses and Disciplines (RFCD) codes.

It was pointed out that the RFCD codes were currently under review, but that any changes would not be made in time for the first round of the RQF.

Participants noted that clarification was needed on whether an output could appear in more than one Evidence Portfolio from the same institution.

The Assessment Process

Participants thought that due to the volume work, and the fact that Panels could not be expected to have the appropriate level of expertise to make judgements on research quality across the entire Panel area; assessors in addition to the Assessment Panel may be required. There were varying views on how this should work, but all participants agreed that any additional assessors would require clear guidance and that the Panel should make the final judgement on a Research Group’s rating.

The question of what would constitute an appropriate volume of material to be sent out to additional assessors was raised with the suggestion that a minimum of 10 research outputs would be required to ensure robust assessment. It was also suggested that at least two assessors should assess each output but it was recognised that this would pose a workload issue.

It was generally agreed by participants that the proportion of outputs sent out to external assessors should be a decision for the Assessment Panels, but that Research Groups should be able to identify, with justification, assessors that they felt were unsuitable for assessing their work.

Outputs for Research Quality Assessment

There was broad agreement amongst participants that there should be no restriction on the type of Research Output put forward in the ‘best four’ (provided it was research) but that some restrictions would apply to the ‘body of work’. It was agreed that guidance should be provided on what Panel Assessors might expect to see in both and that non-standard research outputs should not be disadvantaged.

Participants agreed that where a Group included any non-standard outputs in its ‘best four,’ the onus would be on the Group to provide additional supporting evidence of the quality of the output in the absence of more traditional quality assurance indicators, such as an editorial review process.

Participants agreed that the statement of justification (50 words for traditional and 250 for non-traditional outputs) would be the appropriate place for such claims. These statements could also be used to discuss the Group’s contribution to multi-authored outputs.

Some participants pointed out that the size and scope of a research output varies within and between disciplines; they questioned the minimum best four rule, arguing that some very large outputs should be able to be counted for more than one, for example a multi-
volumed dictionary should be able to be put forward as equivalent to more than one output, or a visual artist may want an exhibition of paintings counted as one output.

**Quality Metrics – Ranked Outputs**

There was general acceptance by participants of ranked outputs as a metric suitable to all Panels but particularly for those Panels where citations were not relevant. A Tiered ranking of ICT conferences, developed by the Computing Research and Education Association of Australasia (CORE) was given as a case study.

The main concerns expressed over ranked outputs included the tight timeframe in which to develop the ranked lists and the difficulty in achieving agreement across all discipline areas. Participants stressed that any exercise to rank research outputs into tiers would need to be coordinated in consultation with the peak bodies representing the various discipline and sub-discipline areas in order to ensure the credibility of this metric.

It was pointed out that some work in ranking outputs into tiers was already underway in the sector in areas such as Education, Economics and Management, Nursing and Midwifery and within the Creative Arts. Participants representing Engineering and Physical Sciences expressed interest in ranking conferences into tiers.

**Quality Metrics – Citations**

Citations were viewed as useful additional information for the majority of discipline areas covered by Panels 1 – 10, provided they were used to supplement rather than replace the peer review component of the process. However, it was generally agreed that citations were of more limited use in the Public Health and many Social Science disciplines. It was agreed that in Panels 11 – 13 standard citations measures were largely irrelevant and should not be applied.

At Workshop 7 a presentation was given on a pilot study undertaken in late 2006 on the usefulness of non-standard bibliometrics for History and Political Science. It was pointed out the while this work required further testing, initial results showed that conventional citation databases could be ‘mined’ for data more relevant to History and Political Science, such as citations to books and other non-indexed publications. As with standard citation measures, it was emphasised that the purpose of this data was to supplement a peer review process.

Participants in all workshops asked that background information on the citation metrics for each discipline needed to be available to the sector prior to institutions preparing their Evidence Portfolios. In particular, institutions needed to know the source of the data, and at least illustrative benchmarks for the discipline-specific centile levels.

There was widespread agreement by participants that Assessment Panels would require specific guidance on how to interpret the citations metric and that they would need to decide how much weight this metric should carry in relation to the other elements of the Evidence Portfolio.

**Quality Metrics – Grant Income**

There was considerable discussion on whether this metric should be expanded to include Categories 2, 3 and 4 funding, as defined by DEST. There was widespread support for Category 1 grant funding being widened to include similar international peer-
reviewed competitive grant schemes. A list of such discipline-specific sources would be sought for inclusion in the panel specifications.

There was no consensus on whether Category 2, 3 & 4 funding was a relevant quality metric. However there was widespread agreement that these could provide quantitative indicators for the assessment of Research Impact.

Rating Research Quality – Interpretation of the Scale

Participants were asked to provide input on a series of issues relating to the rating scale for Research Quality, including the weighting of the elements of the Evidence Portfolio, and whether all elements needed to be rated a five in order for a Research Group to achieve a top rating. Participants were also asked to provide input on the accommodation of equity groups in the rating as well as how international benchmarking or comparisons could be achieved.

Participants agreed that the rating descriptors did not provide adequate guidance in themselves to enable Panels to make a holistic judgement on the overall quality of the group. Participants suggested that more detailed guidance should be given by Panels on how they would interpret the scale and that this should include guidance on weightings and the encouragement of equity groups.

Participants raised the issue of international benchmarking and international recognition as an issue for those disciplines whose research was relevant only within an Australian context. They asked whether in making a case for “world leading” you had to prove that your research was internationally recognised. Participants representing Education pointed out that this would be an issue of particular concern for their discipline. Participants stressed the need for discipline-specific guidance to Assessment panels on assessing the international standing of research.

Participants also remarked that the way the scale is interpreted by the sector would drive the choices that institutions make in identifying Research Groups and eligible researchers for inclusion in the assessment. It was argued that guidance on the interpretation of the rating scale was required in order to ensure that institutions include as many researchers as possible in the assessment.

Domains of Research Impact

Participants across all discipline areas agreed that it would be possible to demonstrate Research Impact across the four domains (social, economic, environmental and/or cultural) but suggested that further guidance was required regarding whether a Research Group had to demonstrate impact across more than one domain to achieve an A rating. Some participants argued that Research Groups should be able to demonstrate impact in as many domains as possible.

Evidence-based Statement of Claims

During workshops participants were asked to identify possible quantitative and qualitative indicators that could be used to support the statement of claims for impact. Some examples of quantitative indicators given included: Linkage Grant Income; repeat business; product sales and web hits. Participants acknowledged that indicators like these were more appropriate for supporting claims of impact at the lower end of the scale, and that evidence supporting higher levels of impact was more likely to be qualitative, i.e. expert testimony.
Participants across all Panel areas agreed that while there should be no restrictions placed on the indicators to support statements of claim for impact, guidance should be given on the types of indicators that would be appropriate. For example, it was suggested by representatives of Engineering that in the case of an engineering group making a claim for impact, Panels would expect to see some indicators that demonstrate engagement with Industry.

The issue of positive versus negative impact was raised at all workshops and participants agreed that all claims for impact should be judged by Panels on their own merits, i.e. how well they demonstrated their case and the strength of the evidence linking the impact to the original research of the Group.

Case Studies

During the workshops there was a wide-ranging discussion over how the case studies that form the bulk of the Impact Statement would be assessed. Some participants saw the case studies as being a maximum of 2 pages each, with the rest of the Impact Statement being used to put forward the context of the Group’s claims for impact.

Participants asked how factors like group size would effect the weighting of the case studies, whether small groups that included fewer than four case studies would be disadvantaged over large groups who included four, and whether it would be necessary for all case studies to illustrate level A Impact for the Group to be awarded an A rating.

Participants in all workshops suggested that these issues should to be trialled, particularly with respect to impact submissions illustrating outstanding benefit.

Participants representing disciplines covered by Panels 9 and 12 argued that guidance should be given in the form of examples; that it would be important to demonstrate the possible variety of claims that could be illustrated by case studies. They also argued for clear guidance to Panels on how to deal with contested research. This was of particular concern for the Humanities and Social Sciences.

Participants also suggested that the importance of research that resulted in levels of impact that demonstrated adoption and engagement should not be undervalued or disadvantaged.

Outputs for Research Impact Assessment

The requirement to demonstrate the evidence linking the statement of claims for impact to the original research of the Group was the area of greatest concern to workshop participants.

Workshop facilitators suggested these concerns could be overcome if the ‘body of work’ had a separate section for listing additional research outputs used to support claims of impact. This solution gained widespread support. It was agreed by participants that there should be no restrictions placed on the type of outputs included in the ‘body of work’ for impact, provided that the output was the result of the original research of the Group.

It was also acknowledged by participants that the outputs listed for impact would not necessarily be the same as that listed for Quality and that the publication window for these outputs included an additional 6 years.
Participants agreed that panel-specific guidance would be required on the types of research output that assessors would expect to see as evidence of underpinning research for impact. Examples included: technical and clinical reports, industry publications, practitioner oriented literature; government reports and submissions to government enquiries, and relevant media coverage.

Some participants expressed concern over how the inclusion of impact outputs like so-called “grey literature” would effect their quality rating, but facilitators pointed out that the Quality Metrics would not be based on outputs listed for impact.

*The Role of End-users*

It was pointed out that what constitutes an end-user or a beneficiary of research would vary across panels and between disciplines and that the wider public would often be the ultimate beneficiaries of research. For some disciplines, end-users would not necessarily be located outside academia, i.e. in the area of Education and Professional Practice.

Participants agreed that the Specifications should include some guidance on the types of end-users that could be contacted to verify a Group’s statement of claims for Impact and that the contact details of end-users could be included in the impact statement but that end-user testimony would not be provided by Groups.

It was agreed that the testimony of end-users could be sought by Assessment Panels where there was some doubt over the link between the Group’s research and the impact being claimed.

*Rating Scale for Research Impact*

Many participants questioned whether you could measure Impact at the same level of granularity as Quality. Participants representing the HASS disciplines commented, with regard to the existing scale, that it was not clear what distinguished the ratings particularly as level C appeared to be of an entirely different order. Participants from the Sciences, however, perceived the scale as progressive.

All participants agreed that discipline-specific guidance was required on what constitutes impact at each level, and that some indication of what a Level A Impact Statement versus a Level C Impact Statement might look like would be useful.

Participants asked how the elements of the Impact Statement would be weighted against the rating scale, and whether impact that was adopted regionally would be disadvantaged against impact that was adopted internationally.

There was general agreement that all of the variables associated with the impact assessment (group size, number of case studies, percentage of researchers engaged in impact related research) would need to be trialled. It was also suggested that during the trial period the outcomes of the various trials run in the sector should be examined more closely with regard to these issues.