

# Research Quality Framework

Assessing the quality and impact  
of research in Australia

**Quality Metrics**

September 2006

At its first meeting on 1 June 2006 the Research Quality Framework Development Advisory Group (RQFDAG) decided to establish a Quality Metrics Working Group (QMWG) to identify the forms and sources of available measures that may assist the process of research assessment by peer review. The QMWG met on 19 June and 21 July 2006 with further investigative work being done by Ms Linda Butler, the Chair. The reports and advice received by the RQFDAG included the following:

## **1. FUNDAMENTAL ISSUES**

Three fundamental issues underpinned the recommendations of the RQF Quality Metrics Working Group (hereafter referred to as QMWG). These were discussed and resolved prior to making final recommendations on the metrics to be used:

### **1.1 The role of metrics**

#### *Recommendation:*

Metrics should be used to inform the peer review process rather than replace it.

This recommendation was elaborated to make the process for the interaction of metrics and the Discipline Panel assessment process more explicit:

- Metrics should not be used in isolation, such as in the 'shadow exercise' proposed for the 2008 UK RAE.
- The data should be made available to Assessment Panels at the beginning of the assessment process.
- No attempt should be made to aggregate indicators to produce a single 'quality score'.

#### *Rationale*

The Working Group saw the role of metrics as enhancing and complementing the Assessment Panel decision-making process. The measures were seen as just one of the elements of the Panel deliberations on quality. There was a strong preference for Panels to have access to the data from the beginning of the assessment process.

It was acknowledged that quantitative measures have the potential to exert undue influence on all Panel decision-making. The QMWG proposed two strategies to lessen these concerns. The first was to recommend that no attempt be made to aggregate the indicators to produce a single score. The second strategy was to propose that the role of the various elements of a Research Group's Evidence Portfolio should be made transparent by stipulating the weight to be given to each in the overall assessment. This proposal is in line with the UK 2008 RAE practice where a minimum weighting for each RAE element is specified (research outputs 50%, research environment 5%, and esteem indicators 5%), but Panels have the flexibility to determine the actual weighting for each element in their discipline within these broad limits. The QMWG left open for further discussion a recommendation on what the three weightings should be in the RQF context.

### **1.2 The number and range of metrics**

#### *Recommendation:*

Assessment Panels should employ a 'basket of quality measures'.

From the list of proposed measures, each Assessment Panel should be free to choose the combination of indicators most appropriate for their disciplines, with some generic measures across all panels.

### *Rationale:*

The character of research 'quality' is complex and multidimensional. No single quantitative measure can address all its facets. The QMWG therefore considered it essential that a 'basket of measures' should be assembled. By not relying on a single metric, the possibility of unintended and undesired responses to the measures would be reduced. The QMWG recommended that some generic measures, applicable to all discipline clusters, should be used to ensure confidence in cross-discipline comparability. However, it was acknowledged that standard citation measures were not applicable in all disciplines, and that an attempt should be made to develop and test alternative equivalent metrics for the applied sciences (particularly computing science), the arts, the humanities and many of the social science disciplines. Possible developments are outlined in more detail in section 3.1 below.

## **1.3 The level of aggregation**

### *Recommendation:*

The measures should be applied to the Research Group's total 'body of work' produced during the RQF census period.

### *Rationale:*

The primary concern behind this recommendation was ensuring that the proposed metrics would be based on a sufficient body of work so as to be robust, thus giving confidence in their use in the RQF process. Concerns were raised in relation to bibliometric measures, particularly those based on citation rates. In any discipline, the distribution of citations can be skewed, with citation rates driven by a few highly cited publications. Preference was stated for bibliometric measures to rest on sets of at least 100 publications. Metrics applied to the 'body of work' of a Research Group, or university output aggregated at the 4 digit RFC code level, would also allow more variation in the measures used.

## **2 THE MEASURES**

### *Recommendation:*

The QMWG proposed three measures of quality (below) to be derived to assist the first round of the RQF. The RQF Development Advisory Group has prioritised the recommended measures as follows:

1. *Ranked outputs:* Distribution of selected category(ies) of research output listed in the Research Group's 'body of work' into ranked tiers, based on predetermined discipline-specific rankings for the selected forms of output. The basis of the rankings will vary by discipline (they could be derived for refereed journals, professional journals, book publishers, conferences, or performance venues). No weightings should be applied to the different tiers;
2. *Citation data:*
  - Citations per publication for the Research Group's 'body of work'; and
  - The proportion of the Research Group's 'body of work' that falls in the top deciles for the discipline (determined from international citation benchmarks);
3. *Grant income data:* Income per Full Time Equivalent for the Research Group based on i) Australian Category 1 sources (as defined by DEST), and ii) International peer reviewed sources. (The DAG view 'grant income' data as a secondary measure).

## *Rationale:*

In deciding on the quantitative measures to be recommended for use in the RQF, the QMWG asserted that the indicators used should: measure some aspect of research quality (and not, for example, refer solely to productivity); be transparent; be reliable when applied to a six year time frame; not involve an excessive financial or time burden to the sector; avoid undue complexity; and should drive desirable behavioural responses from the sector. This last characteristic was regarded as of paramount importance.

### *2.1 Citation data*

The QMWG believed that the overall thrust of the citation measures was to encourage researchers to achieve highly cited publications – but aiming for quality rather than being focussed on quantity – and believed this was a desirable behavioural outcome.

Some concerns were raised that relying solely on ‘citation per publication rates’ as a measure could tempt some Research Groups to limit subsequent output if they had produced a highly cited article, thus enhancing their citation average. To counteract this, a second measure identifying the citation percentile to which each publication belonged (based on discipline-specific yearly citation benchmarks) should be used. While citation rates allow comparison between Research Groups of different sizes, it was also considered essential that the Assessment Panels should be provided with the data that underpinned these averages (i.e. total publication and citation counts, and the number of publications in each percentile).

The citation measures specified are not appropriate in all disciplines. The QMWG strongly supported further investigation into alternative citation metrics for those disciplines where Thomson Scientific databases covered less than half their output. See section 3.1 below for further details on proposed initiatives.

### *2.2 Grant income data*

While some members of the QMWG regarded grant income as an input to research and therefore an indicator of potential only, others supported it as an indicator of ‘quality’ given the assessment of a researcher’s track record is embodied in the grant application process. A decision was made to include it as a metric for the RQF, as it could contribute to the holistic ‘picture’ for the assessors. The QMWG recommended limiting allowable grants to those listed as ‘Category 1’ by DEST, but expanding the coverage to include significant international agencies (e.g. NIH, EU, etc) – the qualifier being that they are international funding sources recognised as using robust merit-based peer review mechanisms. Assessment Panels should decide the international funding sources relevant to their own disciplines.

The QMWG suggested following current HERDC procedures of fractionation in the case of research income received out of a grant spread across more than one institution (e.g. CRC or ARC Centre of Excellence).

The measure was regarded as generic in that it could be applied in all Panel assessments, though benchmarks (ie. average \$/FTE rates) would obviously vary significantly across disciplines.

### *2.3 Ranked outputs*

This indicator is also generic, though its construction would be discipline-specific as the type of output/outlet to be ranked would vary – e.g. journals, conferences, publishers, exhibition venues, etc. The QMWG determined that as the measure is to be used to inform the assessment process, there is no need to weight the tiers and attempt to derive a score for the Research Group.

As with the citation measures, the main thrust of this indicator is to encourage researchers to publish in the most prestigious outlets for their discipline, a response that the QMWG believed was a desirable outcome.

It was noted that many disciplines had already commenced developing output rankings, including computer science, education and the creative arts. However, this work would have to be validated, and other disciplines would need to develop rankings relevant to their own outputs through committees or workshops prior to the RQF process. Discipline peak bodies and the learned academies were identified as potential drivers of these developments.

## 2.4 Measures rejected

*Webmetrics:* At this point measures based on these statistics have not been fully developed as assessment tools. It is anticipated that they may be more robust for future RQF rounds, though there was concern over how readily such data could be audited, and a belief that they might be easily manipulated.

*Collaborations:* with the exception of jointly-authored publications, the effort required to collect data on collaborations is a time consuming process, and its use as a formal metric was rejected.

*Contextual metrics:* The QMWG identified a number of measures that could not be used as stand-alone metrics, but which might be reported in Research Group's context statements. These included measures that related more to a researchers whole career (e.g. esteem measures, service to journals, etc) or were more related to identifying the capacity for generational change (such as research student data and early career researcher numbers).

## 2.5 Who provides the metrics

### *Recommendation:*

Citation analysis should be undertaken centrally where this has been chosen as an appropriate quality measure, similarly ranking analysis, using the 'body of work' provided by the Research Group. Research income and other measures are to be supplied by the Research Groups.

### *Rationale:*

Citation analysis rests on data derived from Thomson Scientific's *Web of Knowledge* indexes. The QMWG noted that universities had varying capacity to undertake in-house analysis of this data. The Australian National University was the only one with access to the raw data files from Thomson Scientific, allowing them to undertake sophisticated analyses. The Group of Eight universities had access to a number of Thomson products that they could use for national and international comparisons, while other universities had more limited options and data sources available to them. If metrics were to be based solely on data available to all universities, their range would be extremely limited. If the RQF sought to use more sophisticated metrics, few universities would have access to the necessary data sources. If universities were given free reign to provide whatever metrics they were able to construct, Panels would likely be faced with a conglomeration of incompatible measures. Therefore the only viable option is for the analysis to be undertaken centrally.

Recent developments from Thomson Scientific will provide the possibility for an efficient and cost-effective method for undertaking the system-wide citation analyses needed for the RQF. This derives from work underway to incorporate an *EndNote* facility, to be known as *EndNote Web*, directly into their *Web of Knowledge* product. *EndNote Web* will be available free of charge to all subscribers to *Web of Knowledge*.

The QMWG discussed the possibilities of using Elsevier Scopus as an alternative, or outsourcing the citation analysis to an overseas institution with the capability to "cleanse" the data, such as

Leiden University. The Group noted that one of the terms of reference was to consider the cost of any measures proposed, and, given that the new developments from Thomson may result in a large cost saving, they were considered a promising solution, and worth investigating further. Other issues that favoured this direction included the limited number of universities accessing Scopus, and its limited acceptance for assessment purposes. The likely outcome is to further examine the *EndNote Web* process.

The QMWG wished to highlight the ease and cost-effectiveness of the *EndNote Web* possibility, though acknowledged that this will require design specifications to accommodate this process in the anticipated RQF Information Management System. The alternative – to extract the citation data by matching submitted publications to raw data files provided by Thomson Scientific – would increase the cost of the exercise significantly, both in money and personnel terms.

For data on other proposed measures – research grant income and ranked outputs – the only possible avenue for collecting the data was for it to be provided by Research Groups. The main stumbling block to obtaining such data from other existing sources was lack of any information disaggregated by RFCD codes. As this type of information was already collected for DEST reporting purposes, the QMWG believed systems were already in place within universities to collect the data. Specific RQF requirements (e.g. a widening of acceptable funding sources, and the classification of outputs into tiers) could readily be incorporated into these systems with some input into the design of the data gathering process.

### **3 FURTHER DEVELOPMENTS**

#### *3.1 Developing and assessing alternative citation metrics*

The use of citation analysis to support the RQF assessment process was supported by the QMWG. There are concerns however, that as a measure for discipline clusters such as the Humanities and Social Sciences, its use cannot be considered unless tested further. The Chief Scientist has met with the Council for the Humanities, Arts and Social Sciences (CHASS) to request that work be undertaken to inform this issue.

Two pilot studies will be undertaken to assess novel citation measures for the disciplines of History and Political Science. This will involve the collection and analysis of 'live' publication data from universities, followed by an assessment of the results by groups of experts in the two disciplines.

#### *3.2 Output rankings*

The project co-ordinated by National ICT Australia and the Australian National University was put forward as 'best-practice' for a process to develop output rankings (in this case, conferences). The project involves extensive initial preparation by a small project team, initial consultations with a broader working group, followed by extensive consultation across the sector. The final stage of the process in this project is to assess the use of measures based on data collected from university ICT departments.

The QMWG expressed the view that the ranking of research output categories to be used as metric indicators in the assessment process for the RQF should be developed prior to the publication of the Discipline-specific Guidelines. Further, it was suggested that this requirement would be assisted by appointing the Assessment Panel chairpersons to oversee this process for each discipline cluster.

## 4 ADDITIONAL ISSUES

### 4.1 Attribution

The QMWG recommended that *multi-authored research* output can be presented in more than one Research Group containing the authors. It was believed this would assist in encouraging collaboration, and would also reduce the complexity of data compilation and analysis. The possibility of tracking multi-authoring of both "intra" and "inter" institutional collaboration would be useful. This could be examined in the design of the Information Management System, although the burden of reporting needs to be minimised.

It would be complex and costly to generate the proposed indicators of quality if, in the event of *personnel movement* in the census period, there was an attempt to attribute research outputs to their source institution or group. The QMWG therefore recommended that all output of the members of a Research Group produced during the RQF census period should be attributed to the group, even if a researcher had moved institutions during the period. It was acknowledged that this could lead to increased staff movements between institutions, but this was not necessarily seen as undesirable behaviour.

### 4.1 Group size

The QMWG chose to discuss the merits of size limits for the proposed Research Groups. For the proposed metrics to be robust (especially citation analysis), the minimum size of a Group should be 5 (assuming this will generate approximately 20 listed items from the 'body of work' per person to be analysed). No comment was made on any maximum size for a Research Group. While the QMWG noted the New Zealand limit of 50 outputs per staff, a number limit for a Research Group's 'body of work' was not supported for the RQF.

---