Donovan, C. (2008). The Australian Research Quality Framework: A live experiment in capturing the social, economic, environmental, and cultural returns of publicly funded research. In C. L. S. Coryn & M. Scriven (Eds.), *Reforming the evaluation of research. New Directions for Evaluation, 118,* 47–60.

4

The Australian Research Quality Framework: A Live Experiment in Capturing the Social, Economic, Environmental, and Cultural Returns of Publicly Funded Research

Claire Donovan

Abstract

The author regards development of Australia's ill-fated Research Quality Framework (RQF) as a "live experiment" in determining the most appropriate approach to evaluating the extra-academic returns, or "impact," of a nation's publicly funded research. The RQF was at the forefront of an international movement toward richer qualitative, contextual approaches that aimed to gauge the wider economic, social, environmental, and cultural benefits of research. Its construction and implementation sent mixed messages and created confusion about what impact is, and how it is best measured, to the extent that this bold live experiment did not come to fruition. © Wiley Periodicals, Inc.

Note: During 2006, the author was chair of the Department of Education, Science, and Training's (DEST) Research Quality Framework Development Advisory Group (RQFDAG) Technical Working Group on Research Impact. The views expressed in this chapter do not necessarily reflect those of the Working Group, the Development Advisory Group, or DEST.

NEW DIRECTIONS FOR EVALUATION, no. 118, Summer 2008 © Wiley Periodicals, Inc.
Published online in Wiley InterScience (www.interscience.wiley.com) • DOI: 10.1002/ev.260

Research funding organizations and science policy circles use the term *research quality* to describe the measurable influence of academic research on the academic community. Research impact denotes the benefits or returns from research, which flow beyond the academic realm to "end users" of research. These end users are traditionally defined as industry, business, government, or more broadly, the taxpayer. As I explain elsewhere (Donovan, 2007b), indicators of research quality such as research income and citation measures have become part of the fabric of research evaluation; in recent years, there has been growing interest in similarly evaluating research impact. This has been spurred by the desire on the part of governments to gauge the value of publicly funded research to end users beyond academia. The reasons for accounting for impact vary: to justify expenditure on academic research in terms of its return on taxpayers' investment, or to create public value for society; to redirect national science foresight planning toward relevant research; to enhance international industrial and economic competitiveness; and (in tandem with quality assessment) to inform performance-based distribution of block funding to universities.

Evolution of Research Impact Evaluation

The following distinct phases in the evolution of impact evaluation have been noted (Donovan, 2007b; Martin, 2007).

Technometrics. The initial search for reliable quantitative measures sought to collate data on investment from industry, commercialization, and technology transfer. However, these data were found to represent a low-order level of impact that did not extend to broader economic or societal benefits, and marked an unsophisticated approach to impact measurement confined to science, technology, engineering and medicine, and the concerns of industry and business (Donovan, 2006).

Sociometrics. A second phase of impact evaluation sought more socially relevant measures in the form of sociometrics, which attempted to map research outcomes onto existing government social statistics (Allen Consulting Group, 2005). Yet, these impact indicators presented no credible link between academic research and macro-level social trends, and they overlooked the cultural import of research.

Case Studies. A third wave of impact evaluations acknowledged that quantification may conceal more than it reveals. Typically employed by dedicated research funding organizations to assess the outcomes of specific funding initiatives, these evaluations proceeded on a case-study basis and sought to combine quality and impact measurements using both quantitative and qualitative (or deliberative) approaches. The case-studies generally attempted to gauge a more broadly conceived notion of impact, which probed various dimensions of the economic, social, and environmental returns from research. This approach demonstrates sensitivity to the definition of *impact,* which varies with the perspectives of such end users as government, citizens,

consumers, business, industry, community groups, NGOs, and practitioners. These different perceptions affect what is valued and hence measured, and so case-study methodology includes several impact dimensions, and it encourages end-user participation throughout the evaluation process (Spaapen, Dijstelbloem, & Wamelink, 2007; Wooding, Hanney, Buxton, & Grant, 2004). Yet, this sophisticated case-study approach has been largely confined to evaluating the impact of scientific and medical research, and it has not been adapted to assessment of a nation's whole research base.

The International Context. In terms of national research assessment exercises, to date the most developed examples of impact evaluation have occurred in the Netherlands and New Zealand. These evaluations sit alongside quality assessment and primarily focus on the economic value of publicly funded research, while measures of broader user engagement are bound to low-level input and output indicators, rather than tangible societal benefits. Even so, the Netherlands seeks data on the influence of research on developments or questions in society at large, and New Zealand collects brief contextual descriptions of linkages beyond academia.

These innovations resonate with current concerns in the international research evaluation community, which has come to recognize the limited value of impact assessment tied exclusively to economic and quantitative concerns; the latest movement in impact evaluation is toward developing richer qualitative and contextual approaches at the national level (Donovan, 2007a; FWF/ESF, 2007). In this vein, the 2008 Australian Research Quality Framework (RQF) was the first national research assessment exercise to include a truly comprehensive and methodologically diverse impact audit.

A Brief History of RQF Development

Australia's RQF came into being as a hybrid solution to academic concerns about research quality and government interest in research impact. The Australian academic community wanted the government to allocate university block funding on the basis of discipline-based peer review of research quality rather than the extant metrics-based formula (DEST, 2004b). The government wanted to boost Australia's innovation strategy through linking academic research to the concerns of industry and business, particularly in the context of broader economic, social, and environmental benefits to society (DEST, 2004a). The RQF was proposed as a panel-based exercise to evaluate both research excellence and the wider benefits of academic research for the nation, and to allocate funds on the basis of outcomes. It is unsurprising, given the impact push from government and the quality pull from academia, that the RQF philosophy of impact evaluation was contested and reshaped throughout its development and implementation.

There were several phases of RQF development involving various advisory groups, technical working groups, and much consultation with the Australian higher education sector. In December 2004, the Minister for

Science, Education, and Training appointed an Expert Advisory Group (EAG), which launched a consultative exercise to determine the structure and features of the RQF (DEST, 2005c). The EAG published its preferred RQF model in September 2005 (DEST, 2005d) and gave its final advice in December 2005 (DEST, 2005b).

A new minister took office, and in March 2006, a new Development Advisory Group (DAG) was created, chaired by Australia's Chief Scientist and tasked to refine the RQF model and detail its phases of implementation. In June 2006, the Minister announced that the first RQF would take place in 2008, and the DAG appointed several technical working groups to address in detail various RQF features in need of further development. This included a Technical Working Group on Research Impact, which reported its findings to the DAG during August 2006.

The Technical Working Group on Research Impact. The Impact Working Group comprised senior academics, senior university managers, representatives from business and industry, experts in impact evaluation, and several DAG members. The membership also represented academic interests in science, technology, engineering, medicine, commerce, humanities, creative and performing arts, and social science. Its remit was to offer detailed advice to the DAG in these areas:

- Methodology: recommend the optimal methodology to assess the impact of Australia's universities
- Indicators: develop generic and discipline-specific quantitative and qualitative measures of research impact
- Assessment period: establish the appropriate length of the assessment period required for effectively assessing research impact
- Evidence portfolios: determine the necessary evidence for research groups to demonstrate impact, including composition of impact statements, metrics to be presented in context statements, and a decision whether "four best outputs per researcher" are adequate to demonstrate research quality and impact
- Demonstrating impact: advise how research groups are to demonstrate research impact, and how ratings of research impact are most effectively reported
- Verifying impact: propose appropriate processes for assessment panels to evaluate research impact

Various RQF features were fixed, and the Impact Working Group had to navigate around them. For example, the EAG had defined research impact as the "social, cultural, economic, and/or environmental outcomes for industry, government and/or other identified communities regionally within Australia, nationally and/or internationally" (DEST, 2005b, p. 24). Another key characteristic was the RQF being a panel-based peer and end-user review of the quality and impact of Australian university research. There

were 13 panels, which were clusters of disciplines that shared similar assessment profiles (for example, physical, chemical and earth sciences; engineering and technology; social science and politics; law, education and professional practices; humanities, creative arts, design and built environment). The assessment will be conducted at the research-group level, rather than at the individual level (as in the case of the New Zealand Performance Based Research Fund) or the discipline level (like the United Kingdom Research Assessment Exercise). The quality assessment consisted of panel judgments combining a peer review of the four "best" outputs per researcher with quality metrics applied to research groups. In terms of impact assessment, set features were an impact scale against which to report and judge the level of research impact; and research groups submitting an impact statement linking the group's research to claimed impact outcomes, the beneficiaries, the measurable difference made by the research, and the details of end users who may confirm research groups' impact claims (DEST, 2005b).

Key Recommendations. The Impact Working Group met four times between June and August 2006. During this period, the DAG offered feedback through its members within the group and by way of DEST, sometimes suggesting that advice be modified—a demonstration of the government pull and academic push in action. At the request of the DAG, the Impact Working Group produced a short outline of its advice, highlighting changes or refinements to EAG recommendations. This was made public as a DAG "Guiding Principles" document in August 2006 (DEST, 2006d). The Impact Working Group presented its final report to the DAG in August 2006 (TWGRI, 2006); the DAG published a revised version of this advice in September 2006 (DEST, 2006b) and its final recommended RQF model in October 2006 (DEST, 2006c). The Impact Working Group's recommendations are summarized below, along with noteworthy deviations from EAG and DAG thinking.

Methodology. The optimum assessment methodology was a qualitative and contextual approach, mediated through the judgment of academic peers and end users. Information is best derived from context statements, impact statements, case studies, and (where appropriate) relevant quantitative and qualitative indicators (TWGRI, 2006).

Indicators. Quantitative metrics are underdeveloped and cannot be used as a proxy for determining impact ratings for research groups; but where appropriate, some qualitative and quantitative indicators may support impact claims (TWGRI, 2006). The DAG decided that assessment panels would nonetheless be given generic impact indicators and be asked to determine additional cluster-specific ones (DEST, 2006c).

Assessment Period. The EAG chose a 6-year assessment window (2001 to 2006) for quality and impact assessment and decided that the impact to be assessed must be related to research conducted within the same 6-year period. The Impact Working Group proposed that although the research impact assessed should occur within the 6-year window, it may be derived

from original research conducted earlier (TWGRI, 2006), and assessment panels may use their judgment to determine a reasonable time frame from the original research to the impact claimed. These recommendations were endorsed by the DAG (DEST, 2006c). The Working Group believed cases where the original research is older than 15 years will require additional supporting evidence (TWGRI, 2006). The DAG limited the period for older research to an additional 6 years only (DEST, 2006c).

Evidence Portfolios. The following recommendations were made in terms of material to be included in evidence portfolios to best enable research groups to demonstrate impact:

1. Impact statements should be the basis of assessing research impact. They should be evidence-based, no more then 10 pages in length, and consist of a statement of claims against impact criteria, up to four case studies illustrating examples of impact, and details of end users who can verify the impact claims (DEST, 2006d).
2. No metrics are to be given in the context statement, but they may be used to support claims made in a research group's impact statement.
3. The EAG proposed the same four best outputs per researcher be used to assess both quality and impact claims. The Impact Working Group recommended that impact assessment should draw on a group's complete body of work, including nontraditional outputs such as reports to government (TWGRI, 2006), which amounted to a revision of the RQF model, a recommendation supported by the DAG (DEST, 2006b).

Demonstrating Impact. The following recommendations were given regarding evidence necessary to demonstrate research impact and how ratings of research impact are most effectively reported:

1. Research impact is best demonstrated by linking a group's impact claims to criteria set out in the impact rating scale. Evidence should connect the group's original research to impact ratings (TWGRI, 2006). The Impact Working Group recommended clear guidelines be developed at the discipline level, a proposal endorsed by the DAG (DEST 2006c).
2. Connecting impact claims to the impact rating scale is the most effective way to report claims of research impact. The EAG recommended a simple three-point scale demonstrating the degree of public benefit derived from research (DEST, 2005b). The Impact Working Group shared this preference, but the DAG directed it to develop a five-point scale with more attention to engagement with end users. The Impact Working Group's final scale was a blend of end-user interaction and public benefit, initially endorsed by the DAG (DEST, 2006d) but later modified by the DAG to reflect more commercial and industrial concerns (DEST, 2006b).

Verifying Impact. Assessment panels will review research groups' evidence portfolios and apply their collective expert judgment to determine the validity of the claims made against the impact criteria. Impact ratings will be assigned, and the rating process will be moderated between discipline panels to ensure consistency and fair treatment for multidisciplinary research. The Working Group recommended the Payback consensus scoring model as particularly suited for this purpose (TWGRI, 2006; Wooding, Hanney, Buxton, & Grant, 2004).

Contested Themes in the RQF Philosophy of Impact Evaluation

According to the Hon. Julie Bishop, Minister for Education, Science, and Training:

> It is my view that if we are able to get right the measure of impact—in both its form and its recognition—then we will have created a research evaluation measure that will greatly surpass those of other nations. (DEST, 2006a)

The role of impact evaluation in the RQF came with high expectations from government. As the minister elaborated, "It will ensure that not only do we, as a country, reward high quality research, but also we reward research which makes a demonstrable change to the way we live or enjoy our lives" (DEST, 2006d). However, the RQF philosophy of impact assessment has, at times, resembled the "pushmi-pullyu" of Dr. Dolittle fame: a two-headed llama that tries to travel in opposite directions. The government push toward impact is offset by a pull toward more scholarly concerns; this push is sometimes forcefully directed toward the interests of industry and commerce, yet counterbalanced by an equally strong pull toward broader public benefits (Donovan 2007c). It is within this context that the chapter now turns to examining central concepts in impact evaluation that display these inherent tensions: defining research impact, communicating research beyond academia, and accounting for research impact.

Defining Research Impact. For RQF purposes, impact was originally concerned with social, economic, and environmental effects, reflecting a trend toward "triple bottom line" accounting (Donovan, 2007b). The EAG's consultation with the higher education sector led to introducing the "cultural" as a fourth impact domain; and the resulting quadruple bottom line was unique in international impact assessment terms. When we turn to consider what, precisely, impact denotes, there are contradictory messages contained in the RQF deliberations, which reflect a fragile balance of push-pull interests.

The Impact Working Group alone supplied actual content for the four impact domains (DEST, 2006b; TWGRI, 2006), which was dropped by the DAG, but reintroduced in the 2007 submission specifications. Impact is

described as adding to the social, economic, natural, and cultural capital of the nation:

- *Social Benefit.* "Improving quality of life; stimulating new approaches to social issues; changes in community attitudes, and influence upon developments or questions in society at large; informed public debate and improved policy-making; enhancing the knowledge and understanding of the nation; improved equity; and improvements in health, safety and security."
- *Economic Benefit.* "Improved productivity; adding to economic growth and wealth creation; enhancing the skills base; increased employment; reduced costs; increased innovation capability and global competitiveness; improvements in service delivery; and unquantified economic returns resulting from social and public policy adjustments."
- *Environmental Benefit.* "Improvements in environment and lifestyle; reduced waste and pollution; improved management of natural resources; reduced consumption of fossil fuels; uptake of recycling techniques; reduced environmental risk; preservation initiatives; conservation of biodiversity; enhancement of ecosystem services; improved plant and animal varieties; and adaptation to climate change."
- *Cultural Benefit.* "Supporting greater understanding of where we have come from, and who and what we are as a nation and society; understanding how we relate to other societies and cultures; stimulating creativity within the community; contributing to cultural preservation and enrichment; and bringing new ideas and new modes of experience to the nation."

We find that the Impact Working Group and early DAG documents define impact in terms of public benefit within these domains (DEST, 2006b, 2006d; TWGRI, 2006). On the other hand, the EAG and the DAG's recommended RQF are concerned with direct practical utility and more targeted groups of end users; for example, *impact* is interchanged with the word *usefulness* (DEST, 2005a) and is "the recognition by qualified end-users that quality research has been successfully applied to achieve social, cultural, economic, and/or environmental outcomes" (DEST, 2005b, p. 12; DEST, 2006c, p. 10); it is found in "short-term . . . outcomes for industry, government and/or other identified communities" (DEST, 2005a, p. 24). We shall see that when considering impact domains and impact rating scales, these divergent views entail mixed messages about what constitutes legitimate impact and how it may be measured and verified.

Finally, when looking at how impact is defined, it is important to note what is excluded. First, the RQF immediately rejected the notion of impact as "knowledge transfer," for example, commercialization of other people's ideas (DEST, 2005d). In this respect, impact may be purely related to a research group's own original research. Second, a research group may apply

for exclusion from impact assessment if its research is at an early stage of development, or if its research orientation means it would be inappropriate to be assessed in terms of impact (DEST, 2005b, 2006b, 2006c). Third, contrary to European developments (FWF/ESF, 2007), basic research is exempted from impact assessment on the grounds that it is not devalued (DEST 2005b, 2006b; TWGRI, 2006): ". . . The fundamental research of today may yield the research impact of the future. In this respect, impact assessment must allow for progress from initial research through to eventual impact, and acknowledge that this is not a necessarily linear process, and that this development takes time" (TWGRI, 2006, p. 2).

However, the "pull" of this sentiment is at odds with the DAG's decision to allow only an additional 6-year window to connect original research to impact; it is a counterintuitive short-term push that devalues basic research through excluding many significant and enduring research impacts from evaluation.

Communicating Research Beyond Academia. During RQF development, there were differing views on what form of publication should be used to link a group's original research to its impact claims. The EAG had recommended that the same best four outputs per researcher be used for both quality and impact assessment (DEST, 2005b). However, this failed to recognize that vehicles for communication differ for academic and nonacademic audiences. It also led to concerns that a linear ideal of scientific discovery underpinned the impact assessment model—that a group of scientists publish a journal article, the idea is taken up and developed, and impact for society is then accrued in terms of technical or health benefits, for example. The Impact Working Group argued that "the types of research output one would submit to demonstrate quality and impact are often quite different because these publications are tailored for different audiences." It recommended that nontraditional outputs, such as reports for government, public exhibitions, and media broadcasts, were an essential link between original research and engagement with end users, and so should be separately drawn on for impact assessment. It also argued that impact occurring within the 6-year assessment period is likely to be connected to traditional and nontraditional research outputs produced before the 6-year window for quality assessment, and thus the window for impact should be extended (TWGRI, 2006). The DAG endorsed these sentiments (DEST, 2006b, 2006c).

In this instance, the push was led by the Impact Working Group's search for the optimum methodology for impact assessment, which polarized the Australian university sector because this preference was supported by technical universities and opposed by the pull of elite academic institutions for an RQF giving primacy to peer review of "high-quality" publications.

Accounting for Research Impact. As has been noted, divergent views of what impact is entail differing views of what should be measured and how. The "push" toward impact as industrial and commercial advance finds

its ultimate expression in quantitative metrics tied to investment from business and industry, patents, and commercialization; the pull toward public value seeks to make previously intangible public benefits of research visible by employing a contextual approach, informed by qualitative and quantitative evidence, and judged by academic peers and end users. Drawing on international best practice in impact evaluation, and strongly favoring a case study approach to methodology, the Impact Working Group supported the latter position. However, we can easily imagine both approaches adopting a panel system informed by evidence supporting a scale of impact claims against the four impact domains presented in the section Defining Research Impact earlier, albeit in a largely quantitative or more contextual manner (potential examples are presented in TWGRI, 2006). Hence, the principle of the case study approach was endorsed by the DAG, as indeed was the continued search (in vain) for more valid metrics of high-order impact (DEST, 2006a).

The impact scale was perhaps the most hotly contested aspect of RQF impact evaluation. As has been noted, during RQF deliberations, the impact scale morphed from a simple three-point measure of degrees of wider benefit, to a fine-grained five-point scale geared to end-user interaction. The actual RQF scale is presented in the next section and matches the DAG preference (DEST, 2006a). It is a linear, progressive scale, premised on a route to impact that begins with (1) engagement with end users who recognize the importance of the research to a defined area, (2) adoption of research, (3) adoption producing benefits for end users, and (4) and (5)evidence of the magnitude of the benefit derived from the adoption. On the other hand, the Impact Working Group's alternative scale was nonlinear, preferring (1) reciprocal engagement with end users, (2) significant uptake of research by the end-user community, (3) significant added value or improvements, and entailing (4 and 5) transformational benefits on a large scale. It was felt that although the language of "adoption" was suited to an idealized model of practice in engineering with industry as the end user, it alienated the humanities, arts, and social sciences. The "pull" was toward a more inclusive scale, which would embrace all disciplines, and the diffuse manner in which research has value beyond academia; the "push" was concerned with targeted end-user engagement and driving behaviors that would make Australia's science base more efficient.

Australia's Live Experiment

RQF development continued throughout 2007, including a series of discipline workshops that each devoted half a day to research impact; further sector consultations; RQF trials, including testing mechanisms for assessing research impact; and development of generic specifications and panel-specific guidance, which were released in September 2007 (DEST, 2007).

The generic specifications display a great deal of pull in that a repeated catchphrase is the usefulness of research for "government, industry, business and the wider community." The definition of impact is extended: "Impact refers to the extent to which research has led successfully to social, economic, environmental, and/or cultural benefits for the wider community, *or an element of the community*" (DEST, 2007, p. 5, emphasis added), which allows inclusion of private value in addition to public value. In terms of defining research impact, there is an explicit request that research groups should include in their impact statements "identifiable and supportable impact-related indicators. This requires the impact statement to identify the beneficiaries of the research and the way in which they have benefited" (DEST, 2007, p. 30). The "push" also dominates in the flavor of examples of impact given: "improved quality of products/services, cost-effectiveness, customer satisfaction, lives saved or productivity" (DEST, 2007, p. 33); "policy impacts can also include changes to policies of corporations, councils, professional groups and non-government organizations" (DEST, 2007, p. 33). A series of examples restricted to industry and psychology are given to illustrate outcomes that would match impact ratings from D to A (see Table 4.1).

The panel-specific guidance, however, does give tailored examples of engagement, uptake of research, and extent of benefit. Yet, no examples of impact metrics are offered. In this sense, the RQF remains a pushmi-pullyu, with the push at the grand policy level and the pull at the research group and panel level, leaving scope for contextual interpretation of the impact scale in discipline-specific terms.

Despite publication of the generic specifications and panel-specific guidance, we find that impact measurement in the RQF remained a live experiment as (1) its fine detail was left to be refined at the panel level,

Table 4.1. The Impact Rating Scale

Rating	*Description*
A	Adoption of the research has produced an outstanding social, economic, environmental, or cultural benefit for the wider community, regionally within Australia, nationally, or internationally.
B	Adoption of the research has produced a significant social, economic, environmental, or cultural benefit for the wider community, regionally within Australia, nationally, or internationally.
C	Research has been adopted to produce new policies, products, attitudes, behaviors, or outlooks in the end-user community.
D	Research has engaged with the end-user community to address a social, economic, environmental, or cultural issue, regionally within Australia, nationally, or internationally.
E	Research has had limited or no identifiable social, economic, environmental, or cultural outcome, regionally within Australia, nationally, or internationally.

although this lack of transparency was of vital concern for research groups in need of guidance to effectively construct their impact statements; and (2) the balance of quantitative indicators versus contextual evidence to inform the second RQF remained under review. The RQF was also a live experiment because of uncertainty over its future; there is a general election due in Australia, and the Labor Party, which was ahead in the polls the day the RQF specifications were released, has vowed that if it replaced the current Liberal coalition government, it would abandon impact assessment. There were suggestions that the RQF would take place in 2009 rather than 2008, and that impact measurement should be a shadow exercise in the RQF's first iteration.

The RQF approach to impact evaluation was a world first; other countries have tended to focus on economic returns or rely on quantitative rather than contextual approaches to impact assessment. The consequence has been that impact measurements prove unsatisfactory, largely because the public value of research has not been adequately addressed. The RQF certainly went a long way toward developing an optimal methodology for capturing the social, economic, environmental, and cultural returns of publicly funded research.

The pushmi-pullyu aspect of implementing a pluralistic impact evaluation may be part of an inevitable compromise of government and academic interests. However, this runs the danger of presenting mixed messages about what precisely research impact is and how best to account for it within a national research assessment exercise. This live experiment did not come to fruition.

Postscript

The Australian Labor Party won the 2007 general election, and on December 21, 2007, announced that the RQF would not proceed. Senator Kim Carr, the Minister for Innovation, Industry, Science and Research, argued that "The RQF is poorly designed, administratively expensive and relies on an 'impact' measure that is unverifiable and ill-defined" (IISR, 2007), The new Government is pursing an alternative evaluation system, Excellence for Research in Australia (ERA). Indications are that this will constitute a "quality" audit without "impact" assessment. In the meantime, impact evaluation is in the ascendancy internationally, and we wait to see if Australia's live experiment will be rekindled, or taken up elsewhere.

References

Allen Consulting Group. (2005). *Measuring the impact of publicly funded research* (Report to the Australian Government Department of Education, Science and Training). Canberra: Allen Consulting Group.

Austrian Science Fund (FWF)/European Science Foundation (ESF). (2007). *Science impact: Rethinking the impact of basic research on society and the economy*. Retrieved September 18, 2007, from http://www.science-impact.ac.at/index.html

Department of Education, Science and Training (DEST). (2004a). *Backing Australia's ability: Building our future through science and innovation.* Canberra: Commonwealth of Australia.

Department of Education, Science and Training (DEST). (2004b). *Evaluation of knowledge and innovation reforms consultation report.* Canberra: Commonwealth of Australia.

Department of Education, Science and Training (DEST). (2005a). *Research quality framework: Assessing the quality and impact of research in Australia—Advanced approaches paper* (Report by the RQF Expert Advisory Group). Canberra: Commonwealth of Australia.

Department of Education, Science and Training (DEST). (2005b). *Research quality framework: Assessing the quality and impact of research in Australia—Final advice on the preferred RQF model* (Report by the RQF expert advisory group). Canberra: Commonwealth of Australia.

Department of Education, Science and Training (DEST). (2005c). *Research quality framework: Assessing the quality and impact of research in Australia—Issues paper* (RQF expert advisory group). Canberra: Commonwealth of Australia.

Department of Education, Science and Training (DEST). (2005d). *Research quality framework: Assessing the quality and impact of research in Australia—The preferred model* (Report by the RQF expert advisory group). Canberra: Commonwealth of Australia.

Department of Education, Science and Training (DEST). (2006a). *Australian government endorses research quality framework* [Media release]. Retrieved September 18, 2007, from http://www.dest.gov.au/Ministers/Media/Bishop/2006/11/B002141106.asp

Department of Education, Science and Training (DEST). (2006b). *Research quality framework: Assessing the quality and impact of research in Australia—Research impact* (Report by the RQF development advisory group). Canberra: Commonwealth of Australia.

Department of Education, Science and Training (DEST). (2006c). *Research quality framework: Assessing the quality and impact of research in Australia—The recommended RQF* (Report by the RQF development advisory group). Canberra: Commonwealth of Australia.

Department of Education, Science and Training (DEST). (2006d). *RQF guiding principles: Research impact* (Report by the RQF development advisory group). Canberra: Commonwealth of Australia.

Department of Education, Science and Training (DEST). (2007). *Research quality framework: Assessing the quality and impact of research in Australia—RQF submission specifications.* Canberra: Commonwealth of Australia.

Department of Innovation, Industry, Science and Research (IISR). (2007). *Cancellation of Research Quality Framework implementation* [media release]. Retrieved April 22, 2008, from http://minister.industry.gov.au/SenatortheHonKimCarr/Pages/CANCELLATIONOFRESEARCHQUALITYFRAMEWORKIMPLEMENTATION.aspx

Donovan, C. (2006, August). An instrument too blunt to judge sharp minds. *Times Higher Education Supplement,* 14.

Donovan, C. (2007a, May). *Accounting for the triple bottom line: A robust qualitative measure of the public value of research.* Paper presented at the Austrian Science Fund/European Science Foundation Conference on Science Impact: Rethinking the Impact of Basic Research on Society and the Economy, Vienna.

Donovan, C. (2007b). The qualitative future of research evaluation. *Science and Public Policy, 34*(8), 565–574.

Donovan, C. (2007c) Introduction: Future pathways for science policy and research assessment: Metrics vs. peer review, quality vs. impact. *Science and Public Policy, 34*(8), 538–542.

Martin, B. (2007, May). *Assessing the impact of basic research on society and the economy.* Paper presented at the Austrian Science Fund/European Science Foundation Conference on Science Impact: Rethinking the Impact of Basic Research on Society and the Economy, Vienna.

Research Quality Framework Technical Working Group on Research Impact (TWGRI). (2006). *Final report: Optimal methodology for assessing research impact.* Canberra: TWGRI.

Spaapen, J., Dijstelbloem, H., & Wamelink, F. (2007). *Evaluating research in context: A method for comprehensive assessment.* Hague, Netherlands: COS.

Wooding, S., Hanney, S., Buxton, M., & Grant, J. (2004). *The returns from arthritis research, vol. 1: Approach, analysis and recommendations.* Cambridge: RAND Europe.

CLAIRE DONOVAN is a research fellow at the Research Evaluation and Policy Project in the Political Science Program, Research School of Social Sciences, the Australian National University.