This report is part of the RAND Corporation research report series. RAND reports present research findings and objective analysis that address the challenges facing the public and private sectors. All RAND reports undergo rigorous peer review to ensure high standards for research quality and objectivity.

Assessing Research Impact

An international review of the Excellence in Innovation for Australia Trial

Molly Morgan Jones, Sophie Castle-Clarke, Catriona Manville,
Salil Gunashekar, Jonathan Grant

RAND EUROPE

RAND EUROPE

Assessing Research Impact

An international review of the Excellence in Innovation for Australia Trial

Molly Morgan Jones, Sophie Castle-Clarke, Catriona Manville,
Salil Gunashekar, Jonathan Grant

Prepared for the Australian Technology Network of Universities

The research described in this document was prepared for the Australian Technology Network of Universities.

Published 2013 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
Westbrook Centre, Milton Road, Cambridge CB4 1YG, United Kingdom
RAND URL: http://www.rand.org
RAND Europe URL: http://www.rand.org/randeurope
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Preface

The Australian Technology Network of Universities asked RAND Europe to review the Excellence in Innovation for Australia (EIA) Impact Assessment Trial ('the EIA Trial'), in order to assess how well universities identified and demonstrated impact, as well as how the process could be further improved.

This report offers headlines regarding the success of the process, as well as actionable recommendations for improving the EIA Trial in its current form, and for scaling up the process in the future. It also includes a detailed review of the Trial guidance, an analysis of case studies submitted to the Trial, an analysis of how each case study was scored by the assessment panels and an analysis of surveys completed by institutions and case study authors. The survey was conducted in January 2013.

The report is intended for those responsible for the EIA Trial, in order to enable them to improve the exercise. However, it may also be of interest to others working in the evaluation of research impact.

RAND Europe is an independent not-for-profit policy research organisation that aims to improve policy and decisionmaking in the public interest, through research and analysis. RAND Europe's clients include European governments, institutions, non-governmental organisations and firms with a need for rigorous, independent, multi-disciplinary analysis. This report has been peer-reviewed in accordance with RAND's quality assurance standards.

For more information about RAND Europe or this document, please contact:

Jonathan Grant
RAND Europe
Westbrook Centre
Milton Road
Cambridge CB4 1YG
United Kingdom
Tel. +44 (1223) 353 329
jgrant@rand.org

Contents

Table of figures and tables

FIGURES

TABLES

Executive summary

This report reviews the process of the Excellence in Innovation for Australia (EIA) Impact Assessment Trial ('the EIA Trial'). The EIA Trial had two aims: to assess the non-academic impact of research generated by a subset of Australian universities, and to be a pilot for a potential companion piece to the next Excellence in Research for Australia (ERA), a nationwide performance assessment of Australian universities.

Twelve universities participated in the Trial, including the Australian Technology Network of Universities (ATN), which commissioned this review. The participating universities submitted a total of 162[1] research impact case studies, which were assessed by seven independent panels, made up of 75 experts, of whom 70% were from outside the university sector. The results of the Trial were published in November 2012.

The aim of this review was to provide an overall assessment on the applicability of the EIA process as a companion piece alongside future rounds of ERA. The review had two objectives: to evaluate the process to see how EIA could be improved if repeated in the future, and to evaluate how the EIA could be scaled up to be a companion piece to the ERA.

The review resulted in eight headline observations that are summarised in Box 1 and described in more detail below. The headlines were derived by combining RAND Europe's expertise in research impact assessment, together with the synthesis of four evidence streams collected for this study:

- *A critical review of the assessment guidelines and other relevant documentation*. The guidelines published for the Trial were benchmarked against other international impact assessment exercises including the UK Research Excellence Framework (REF).
- *A critical review of a sample of submitted case studies*. Ten case studies, sampled from ATN universities and deliberately chosen to ensure a spread of subject areas and ratings, were reviewed to assess the quality of case studies, the application of the guidelines and how panels assessed the case studies.

[1] Of the 162 unique case studies that were submitted, one was sent to two socio-economic objective (SEO) sectors for assessment and another to three SEO sectors, resulting in an aggregate of 165 case studies by SEO sector.

- *An assessment of panel scoring*. An institutionally de-identified quantitative assessment of the overall ratings of the four main panels (including sub-panels) to assess scoring behaviour, including inter- and intra-panel reliability.
- *Surveys of ATN universities*. A survey of the five ATN member universities to provide an institutional perspective of the Trial, and a survey of case study authors to get a researcher perspective. All five institutions responded, and 24 of the 64 individual case study authors responded.

The eight headlines are divided into two groups. The first focus on improving the EIA if it was to be repeated; the second focus on issues and ideas that would need to be considered if the EIA was to be scaled up as a companion piece for the ERA. Neither set of headlines are necessarily mutually exclusive, and all eight should provide the basis for continued discussions and consultations about how to develop a solid foundation for impact assessment within the Australian university sector.

Box 1: Summary of headline findings from review of the EIA Trial

Issues and ideas on how the EIA could be improved if repeated

1. The EIA Trial was a successful process, but it could be improved
2. The quality of case studies could be improved
3. Consider including the contribution of research as an assessment criterion (in addition to reach and significance)
4. There is a need, and an opportunity, to create and embed a culture of and expectation for impact within Australian universities and wider society
5. Steps should be taken to understand the benefits and (unintended) consequences of future assessments

Issues and ideas on how the EIA could be scaled up as a companion piece to the ERA

6. The EIA impact assessment process will need to be reconfigured if scaled up
7. Any future EIA impact assessment will need to ensure that the direct and indirect benefit of assessing impact in Australian universities outweighs the transaction costs of the assessment
8. There will be a need to engage early with other institutions that did not participate in the EIA Trial in a scaled up impact assessment

1. The EIA Impact Assessment Trial was a successful process, but it could be improved

The Trial was a success and met its aims and objectives. It is a unique exercise as it is one of the largest attempts worldwide to assess the non-academic impact of research arising from the higher education sectors using a case study methodology. In doing this, it encouraged reflection on the non-academic impact of Australian universities, allowed participant institutions to demonstrate their impact, and demonstrated the feasibility of assessing research impact through a case study approach. The survey responses showed that institutions and case study authors were positive about the Trial, the review of the case studies demonstrated engagement in the process and also that impact could be identified and articulated, and the assessment of the panel scores showed how panels could differentiate between different degrees of impact. The Trial also identified areas for improvement in the process for assessing impact, as detailed below.

2. The quality of case studies could be improved

For the assessment of impact to be fair, assessed differences between case studies must be reflected in the relative reach and significance of the impacts, and any artefacts of the drafting process must be minimised. The scoring analysis showed that the individual intermediate panel scores about the description and nature of the impact were the most significantly correlated with the overall case study score, as compared to scores about the description of how the research underpinned the impact and/or the validation of the impact. With this in mind, there is an opportunity to improve the quality of the case studies by making some changes to the guidance documentation, by reordering the case study template, and by providing material and training on how to write good impact case studies.

The guidance material provided to authors and panels was very concise and clear and was praised by all the institutions in the survey. However all institutions also provided suggests for how the guidance could be improved (as described in section 2.2.1). Two of five surveyed institutions specifically noted that reordering the impact case study template would significantly improve the overall quality of the case study and another noted the chance of repetition within the existing sections. The provision of best case examples was suggested by the majority of institutions and noted by a number of authors. The benefit of such material was reinforced from the review of the case studies, which highlighted a number of technical 'errors' in articulating impact (e.g. focusing on potential not realised impacts) that were overlooked by case study authors and in some cases by the panel. We suggest that it may be beneficial to provide best-practice examples of impact case studies or practical advice to case study authors in future, and consider publishing impact case studies from the trial, and their scores, if possible, although we recognise that doing this will not be straightforward.

3. Consider including the contribution of research as an assessment criterion (in addition to reach and significance)

There are two main criteria in the EIA guidelines for assessing research impact: reach and significance, where reach refers to the spread or breadth of influence or effect on the

relevant constituencies, and significance refers to the intensity of the influence or effect. However, the EIA guidance does not align reach and significance with contribution for either contribution of the research to the impact, or the contribution of various research groups to the overall research project (and hence the impact). This is important as the impacts claimed in a case study could be large, but the contribution of the claiming researcher or research group could be relatively small, and this needs to be taken into account when assessing impact. It was notable that none of the case studies reviewed attempted to articulate how research contributed to the impact and indeed this issue did not arise in either survey. Thus it may be beneficial to amend the guidance for case study authors and panel members in order explicitly to address how contribution of the research to the impact is accounted for and to include this as a criterion alongside research and significance in future rounds of EIA.

4. There is a need, and an opportunity, to create and embed a culture of and expectation for impact within Australian universities and wider society

Australia has been at the forefront of developing national research evaluation and impact assessment systems over the past ten years. The Research Quality Framework, which was subsequently superseded by ERA, originally aimed to assess the quality and impact of research. This suggests that the concept of researcher accountability, and accountabilities that go beyond the academic systems, are relatively well founded within the Australian higher education system. This was confirmed in the institutional survey: three of the five respondents said the Trial demonstrated the need to show impact regularly. It will be important for the sector to continue to create and embed a culture of impact within Australian universities and also to create the demand for and expectation of such an impact by the tax paying public and their representatives. Institutions will need to develop internal impact strategies that focus on not only institutional approaches to the information management of impact, but also how to maximise impact as and when it occurs.

5. Steps should be taken to understand the benefits and (unintended) consequences of future assessments

A number of practical issues come to the fore when future rounds of EIA are considered. For example will a future assessment look at the marginal improvements in impact – the additional impact that has accrued between rounds – or allow the total impact to be (re)assessed? Is there a risk that institutions and individuals will attempt to manage the timing of impact to fit assessment windows? Will wider stakeholder engagement – both panel participation and support for the 'evidencing' of impact – hold up? The answers to these types of questions will need to be discussed and debated prior to subsequent iterations of EIA. Regardless it will be important that clear, transparent guidance be issued as soon as feasible if an EIA exercise is to be repeated, as that guidance will help to prevent unwanted or unnecessary behaviours.

6. The EIA impact assessment process will need to be reconfigured if scaled up

If the EIA Trial is to be scaled up as a companion piece to ERA it will need reconfiguration because the primary purpose of assessment will need to be clarified and may change; the unit of analysis may be too broad at a larger scale; and there is a limit to the number of case studies that can feasibly be reviewed by panels.

First, there will be a need to seek agreement on the primary purpose of the assessment as this will influence design. For example, will a scaled up EIA lead to the allocation of funding to Australian universities? The strategic intent of the exercise may influence the number of case studies submitted, the unit of analysis, and the nature of engagement of future design considerations.

Second, there may need to (re)consider the unit of analysis. A unique characteristic of EIA was its focus on socio-economic objective (SEO) codes as the organising principal, as opposed to research discipline (as adopted by the UK REF) or fields of research (as in ERA). Although we understand the rationale for using SEO codes and the merit in doing so as it relates to the underlying purpose of the EIA and the broader context in Australia, conceptually this may have the potential of limiting the variety of impacts that can be assessed, and which might be raised by case study authors. Although this is analogous to debates about how to deal with multi-disciplinary research when the organisational principal is disciplinary focused, it is nevertheless worth considering the implications. Two of the five institutions surveyed thought the SEO codes were an appropriate unit of analysis and another stated that groupings were acceptable, but three institutions noted the main sectoral panels under which the SEO codes were grouped (defence, society, economic and environmental) were too broad and did not allow for the benefits of the disaggregation by SEO code to be reflected as strongly as they could be. We suggest a more disaggregated system would indeed be needed with larger numbers of participating institutions, cases studies and panels. Institutions also noted that more guidance would be needed on how to deal with case studies that cross a number of SEO codes.

Finally, future exercises will need to consider the number of case studies that can feasibly be assessed. This will be a trade-off between generating enough case studies for a representative assessment of an institution's impact, and the workload (and costs) of institutions, panels and administrators. These and other issues are interdependent, and thus likely to be resolved through an iterative process of development.

7. Any future EIA will need to ensure that the direct and indirect benefit of assessing impact in Australian universities outweighs the transaction costs of the assessment

Any scaled up impact assessment system needs to provide value for money. Based on typical transaction costs for research funding, the cost of funding the impact assessment should be no more than a tenth of the benefits to participating institutions. The costs associated with the identification, articulation and evaluation of impact case studies are estimated to be about five to ten days (or AU$5k-AU$10k) per case study. This is very similar to time estimates generated for the pilot of the UK's REF. If a scaled up EIA is to aid the allocation of funding alongside the ERA in the future, it is possible to project the most logical level of funding allocation that should result from the impact assessment.

Based on estimated calculations presented in this report, we suggest that for every case study submitted for assessment, about AU$100,000 would need to be available for allocation to universities to make the exercise worthwhile. Combining this heuristic with a discussion about the design of a scaled up system will help to clarify and test some of the trade-offs described above (headline 6).

8. There will be a need to engage early with other institutions that did not participate in the EIA Trial in a scaled up impact assessment

The EIA Trial involved 12 of Australia's 39 universities and thus it will be essential to involve other institutions in the development of a scaled up version of EIA. This will help to generate sectoral ownership for a national impact assessment system and to ensure that non-participating institutions from the EIA Trial are not at a disadvantage. Related to this, and as already discussed, will be the need to provide material such as examples of case studies and 'how to' guides, as well as brokering and facilitating a sector-wide discussion and debate. At the same time it will be important not to dilute some of the unique strengths of EIA including, for example, its focus on SEO codes, the dominance of non-academic panel members, and the brevity and clarity of guidance.

Acknowledgements

RAND Europe would like to thank Vicki Thomson and Dr Matt Brown at the Australian Technology Network of Universities for their guidance throughout this project. We are also very grateful to all case study author and institutional survey respondents from Curtin University, Queensland University of Technology (QUT), Royal Melbourne Institute of Technology (RMIT), University of South Australia (UniSA) and University of Technology, Sydney (UTS). We would also like to thank Sunil Patel for his very helpful insights and analyses for some of the panel scoring assessment calculations.

We thank our Quality Assurance reviewers Michael Frearson and Linda Butler for their helpful comments on the report.

CHAPTER 1 Introduction

1.1 The origins and aim of this report

This report reviews the process of the Excellence in Innovation for Australia (EIA) Impact Assessment Trial ('the EIA Trial'). The EIA Trial aimed to assess the non-academic impact of research generated by a subset of Australian universities and to be a pilot for a potential companion piece to the next Excellence in Research for Australia (ERA) – a nationwide performance assessment of Australian universities. Twelve universities participated in the EIA Trial, including the Australian Technology Network of Universities (ATN), which commissioned this review. The participating universities submitted a total of 165 research impact case studies that were assessed by seven independent panels, made up of 75 experts, of whom 70% were from outside the university sector. The results of the EIA Trial were published in November 2012[2] and are summarised in Box 2.

Box 2: Summary of EIA Trial key conclusions

1. It is possible to assess research impact across a broad range of disciplines
2. The case study approach can provide a compelling narrative of the impact of research
3. Research impact could be assessed against an outcomes based system of classification such as the SEO codes, recognising that there are some limitations to this methodology
4. Expert panels comprising a majority of end-user stakeholders are able to assess research impact; the panels should also include an appropriate discipline mix covering the breadth of research impacts being considered
5. Development of an impact component of any broader research assessment exercise would require further consideration of the number of case studies to be submitted

[2] Australian Technology Network of Universities and Group of Eight (2012) *Excellence in Innovation Research Impacting Our Nation's Future – Assessing the Benefits*, Australian Technology Network of Universities.

The aim of this review is to provide an overall assessment on the applicability of the EIA process as a companion piece in future rounds of the ERA. Specifically this includes evaluating the process developed for the EIA and considering what issues might arise if the process is scaled up. To do this we synthesised information across four evidence streams:[3]

- *A critical review of the assessment guidelines and other relevant documentation*. The guidelines published for the Trial were benchmarked against other international impact assessment exercises including the UK Research Excellence Framework (REF).
- *A critical review of a sample of submitted case studies*. Ten case studies, sampled from ATN universities and deliberately chosen to ensure a spread of subject areas and ratings, were reviewed to assess the quality of case studies, the application of the guidelines and how panels assessed the case studies.
- *An assessment of panel scoring*. There was an institutionally de-identified quantitative assessment of the overall ratings of the four main panels (including sub-panels) to assess scoring behaviour, including inter- and intra-panel reliability.
- *Surveys of ATN universities*. There were surveys of the five ATN member universities to provide an institutional perspective of the Trial, and of case study authors to get a researcher perspective. All five institutions responded, and 24 of the 64 individual case study authors responded.

In addition, we bought to bear RAND Europe's experience of research evaluation and impact assessment,[4] including precursor studies for the UK REF[5] and ongoing reviews of research case studies for UK universities as part of their preparation for REF.

1.2 The unique characteristics of the EIA Trial

The EIA Trial is one of the largest attempts worldwide to assess the non-academic impact of research arising from the higher education sectors using a case study methodology. The only comparator exercise is the UK REF impact pilot, which involved 29 institutions submitting 324 case studies for review.[6] However, there are some fundamentally important differences between the UK REF impact pilot (and its implementation) and the EIA Trial that need to be stressed in the context of this review. These innovations include the dominance of industry involvement in the expert review panels and the centrality of impact as the unit of evaluation. As noted above, about 70% of the EIA Trial panel

[3] It should be noted that it was outside the scope of this project to review any direct feedback from assessment panel members on the EIA Trial either by interview or survey, particularly industry based panellists.

[4] Guthrie, S. et al. (forthcoming) *Measuring Research: A Guide to Research Evaluation Frameworks and Tools*, Santa Monica, CA: RAND Corporation.

[5] Grant, J. et al. (2010) *Capturing Research Impacts: A Review of International Practice*, Santa Monica, CA: RAND Corporation.

[6] Technopolis (2010) *REF Research Impact Pilot Exercise Lessons-Learned Project: Feedback on Pilot Submissions: Final Report*, Technopolis, November 2010.

members came from outside the university sector; by comparison currently 12% of appointed REF panels are users, although during 2013 it is expected that additional appointments will be made. The other difference is that the unit for assessing impact in the EIA Trial is defined by social and economic objectives rather than the research underpinning the impact, as with REF.

1.3 Why assessing research impact matters

The assessment of the non-academic impact of research is not new,[7] but there is a growing interest internationally in methodological approaches to the measurement of research impacts.[8] Here we take research impacts to be a demonstrable contribution to something outside the academic system. The increased interest in measuring research impact is due to a number of different, but not mutually exclusive, drivers: advocacy, accountability, analysis and allocation. Each driver has a slightly different rationale, with corresponding implications for how impact might be evaluated and presented.

First, set against a background of fiscal austerity and a drive for efficiencies in many nations, research funders and providers are having to compete with other public services in *advocating* the need for, and continued funding of, research. Leaders within the sector need to have compelling arguments to 'make the case' for research in such discussions. These arguments are often best made at the macroeconomic level, supported by compelling narratives or case studies that are representative of research impact. Here, one could demonstrate a comparatively attractive rate of return for research investments or a significant cost in not making an investment, which becomes more tangible when supported by a narrative that is illustrated at a micro level through case studies.

Related to this is the *accountability* of the research community to those who fund its activity, be they tax payers or donors. Good governance dictates that the recipients of public funding should be able to provide an account of their decisionmaking. In the context of research and higher education funding, we would argue that this means that sound grant-making decisions are made in a transparent, merit-based way that lead to some public benefit or social impact beyond academia. Such accountability can occur through the collection and reporting of appropriate metrics of activity, output, outcome and impacts as illustrated by the UK Medical Research Council's E-Val system.[9]

Another reason for assessing research impact is to provide what is analogous to a dependent variable in the *analysis* of research policy or, put another way, to understand what works in research funding. The 'science of science', as it is has become known, is predicated on the ability to measure research with the aim of improving the effectiveness and value for

[7] Marjanovic, S., Hanney, S. and Wooding, S. (2009) *A Historical Reflection on Research Evaluation Studies, Their Recurrent Themes and Challenges*. Cambridge, UK: RAND Europe.

[8] OECD (2010) *OECD-Norway Workshop on Performance-based Funding for Public Research in Tertiary Education Institutions*; Academy of Medical Sciences (2006) *UK Evaluation Forum Medical Research: Assessing the benefits to society*, London: Academy of Medical Sciences.

[9] Medical Research Council (2010) *Outputs, Outcomes and Impact of MRC Research: Analysis of MRC e-Val Data*.

money for research funding.[10] Or as John Marburger, the former Science Advisor to President George W. Bush, put it in an editorial in *Science* in 2005: 'A new "science of science policy" is emerging… But [this] demands the attention of a specialist scholarly community. As more economists and social scientists turn to these issues, the effectiveness of science policy will grow, and of science advocacy too.'[11] Therefore, knowing 'what works' and why will require consideration of which areas of science to invest in, determination of how and who should invest, and identification of the returns.[12]

Finally, the *allocation* of research funding based on non-academic impact is relatively new, the UK REF being the first time it has applied across a research system. In November 2013, UK universities will make REF submissions. The REF will assess universities on the basis of the *quality* of research outputs, the *wider impact* of research and the *vitality* of the research environment. As with the EIA Trial, following a pilot exercise,[13] the UK funding councils concluded that the peer review of research impact case studies is a workable approach. The weighting for the impact assessment part of REF will be 20% of the total assessment in 2014, and this is likely to rise to 25% in the future. Even at 20%, this equates to around £220m (AU$326m) per year, and so constitutes a significant amount of funding for research.

In the context of the EIA Trial it is important to note that the main focus is on accountability and advocacy, but one of the questions of this review was to see whether it can be applied also for allocation purposes, as is planned for the UK REF. It is also worth noting that the curation of a library of research impact case studies will support analysis into factors associated with successful research translation. However, irrespective of the primary purpose of the impact assessment is the need to understand, acknowledge and where feasible address some of the perennial themes and challenges that make assessing research impact difficult.

1.4 Why assessing research impact is difficult, but not impossible

There are a number of challenges that need either to be acknowledged or, where possible, addressed in the development of a system or approach to measure research impact. These challenges are not new[14] and of differential importance depending on the primary purpose of the assessment. They also provide some criteria to assess the EIA Trial as set out later in this report. Below we identify four key challenges.

---

[10] Grant, J. and Wooding, S. (2010) *In Search of the Holy Grail: Understanding Research Success*, RAND_OP295, Cambridge: RAND Europe.

[11] Marburger III, J.H. (2005) 'Wanted: better benchmarks', editorial, *Science* 308(5725): 1087.

[12] Chataway, J. and Morgan Jones, M. (2011) 'Analyse this', *Public Service Review: European Union* 21, pp. 324–6; Grant, J. and Wooding, S. (2010) *In Search of the Holy Grail: Understanding Research Success*, RAND_OP295, Cambridge: RAND Europe.

[13] Technopolis (2010) *REF Research Impact Pilot Exercise Lessons-Learned Project: Feedback on Pilot Submissions: Final Report*, Technopolis, November 2010.

[14] Academy of Medical Sciences (2006) *UK Evaluation Forum Medical Research: Assessing the benefits to society*, London: Academy of Medical Sciences.

1.4.1 Time lags

The time it takes for research to translate from academia into wider societal benefits is largely unknown, and where known it is highly variable, but in the biomedical and health sciences has been estimated to be on average 17 years in a number of different studies.[15] This means that any assessment of contemporary impact may have to look at research that could have occurred two decades ago. Within the context of impact assessment this raises a number of other challenges. First, is it possible to recall such research with appropriate accuracy? Second, is the context within which the research occurred relevant today? If not, how does this impact on our understanding of the wider research system and our ability to analyse factors associated with successful translation of research; and is it fair to reward and allocate research funds based on such historical performance? Finally, given the mobility of university-based researchers, is it possible accurately to attribute research activity to one institution given that a researcher and research team may work in many different universities across a career?

1.4.2 Contribution and attribution

The linkage between input, activity, output and outcome is central to developing an understanding of how research translates into public benefit. However, in research impact assessment this can prove difficult to determine and evidence. At its simplest, a research project is likely to have multiple research inputs – different research grants, collaborators working in different research institutions, researchers at different stages of a research career – and result in multiple research papers (outputs). These multiple linkages get increasingly complex as one progresses downstream to assess research outcomes and non-academic impacts. The challenge, then, of any system that evaluates research impact is to ensure that we have an understanding of the 'contribution' and 'attribution' relative to the outputs, outcomes or impacts that result from the research input and activity. Here, contribution refers to the relative efforts made by a research team(s) and the relationship to the outputs, outcomes and impacts, whereby attribution refers to the proportional extent to which the outcomes or impacts have resulted from those efforts and research outputs.[16] The way contribution and attribution are, or should be, highlighted will be differentially important depending on the purpose of the assessment. If the purpose is advocacy or accountability, practitioners are increasingly relying on contribution.[17] However, if the purpose is analysis or allocation it is important to attempt to take into account the relative contribution and attribution of the research on the impact. Assessing this is complex, however. At its theoretical extreme the assessors need to ask the counterfactual question: 'What would have happened if the research did not occur?'[18] Ideally a baseline assessment of what the

[15] Slote Morris, Z., Wooding, S. and Grant, J. (2011) 'The answer is 17 years, what is the question: understanding time lags in medical research', *Journal of the Royal Society of Medicine*, 104(12), pp. 510–20.

[16] This distinction is a subtle, yet important one that is found most often in the evaluation literature. Here, we acknowledge its importance, yet throughout this report we refer primarily to contribution when considering the relationship between research and impact as this is the language used in the EIA guidance documents.

[17] Canadian Academy of Health Sciences (2009) *Making an Impact: A Preferred Framework and Indicators to Measure Returns on Investment in Health Research. Report of the Panel on the Return on Investments in Health Research*.

[18] Sometimes the counterfactual is also framed in the concept of 'additionality', which is what is the additional benefit of the intervention (or research) over the baseline.

status quo would have looked like without the research is compared to the world with the research. In practice, however, research is often incremental, collaborative and iterative, so isolating the contribution and attribution of a particular piece of research to a given set of outputs, outcomes and impacts is challenging and inevitably relies on some form of judgement.

1.4.3 Assessment at the margin

Any measurement system must be able to differentiate and scale different research impacts. In bibliometric analysis relative citation counts provide an easy but not uncontroversial way for scaling the impact of different papers, but this gets more challenging when using more judgement-based approaches such as the review of case studies. Is it possible, for example, for experts to differentiate between a high and low research impact case study? What is it that marks the boundary between high and low impact? Do experts agree on the marginal differences and if so would that agreement be accepted by the wider research community? Clearly there is an analogue with the allocation of research grants based on peer review or assessing undergraduate degrees where academics are well practised at differentiating between a second class honours and a first class degree.

1.4.4 Transaction costs

All forms of research assessment are costly. We know, for example, that the Higher Education Funding Council for England spent £5.6m running the 2001 Research Assessment Exercise in the UK (less than 1% of the total funding budget) and the total administrative cost of the funds provided by UK research councils to universities is less than 10% of the funds awarded.[19] In both cases the assessment has been focused on research quality and, given the longevity of both systems, has been largely accepted by policy and academic communities (with some dissent), suggesting that benefit outweighs costs. However, the introduction of impact assessment alongside the traditional assessment of research quality inevitably adds to the transaction costs. The question for policymakers is whether the anticipated additional benefits of such additional costs are justifiable (or, as originally proposed for REF, whether the costs of assessing research quality could, for example, be reduced through the widespread adoption of metrics).

1.5 Report structure

The remainder of this report is organised around the eight headline issues that were identified over the course of our review. These are divided into two groups: the first set of five headlines address issues and ideas for improving the EIA Trial process if it were to be repeated again; the second set of three headlines addresses issues and ideas that need to be considered if the EIA Trial is to be scaled up and used as a companion piece to the ERA. Though presented separately, both sets of headlines are not mutually exclusive. We draw throughout the report on data and information from the review's four main workstreams. Full accounts of the analyses from each workstream are provided in the annexes. Annex A discusses the critical review of the EIA guidelines and other impact evaluation frameworks

[19] Sastry, T. and Bekhradnia, B. (2006) *Using Metrics to Allocate Research Funds: A Short Evaluation of Alternatives to the Research Assessment Exercise*, Higher Education Policy Institute; RCUK (2006) *Report of the Research Councils UK Efficiency and Effectiveness of Peer Review Project*, Swindon: Research Councils UK.

and documents. Annex B presents the findings from the critical review of the selection of ATN case studies. The panel scoring assessment is provided in Annex C and the survey findings and analysis is in Annex D. We strongly recommend that the report is read together with the annexes in order to appreciate the full extent of the analysis.

CHAPTER 2 Issues and ideas for improving the EIA Trial

In this chapter we set out the five headline observations focused on improving the EIA if it was to be repeated. Chapter 3 describes the issues and ideas that would need to be considered if the EIA was to be scaled up as a companion piece for the ERA. Neither set of headlines is necessarily mutually exclusive, and all eight should provide the basis for continued discussions and consultations about how to develop a solid foundation for impact assessment within the Australian university sector.

2.1 The EIA Trial was a successful process, but it could be improved

The objectives of the EIA trial were to 'a) demonstrate the innovation dividend of research generated by Australian universities and, b) be a precursor to a possible companion piece to Excellence in Research for Australia (ERA) in the allocation of research funding'.[20] In light of this, we conclude on the basis of our review that the EIA Trial was a successful process and we agree with the statement in the foreword to the EIA report made by Mr Philip Clark that 'the EIA trial has clearly demonstrated that Australian universities can generate compelling case studies of impact across a wide range of disciplines and impact areas'.

We make this statement on the basis that the EIA Trial encouraged reflection on the wider impact of the Australian universities that participated, and allowed them to demonstrate their impact. The success of the EIA Trial also lies in the fact that the impact of university research was able to be communicated and articulated in case studies. In other words, universities were not only able to identify impact, but the whole process also demonstrated the feasibility of identifying, assessing and making distinctions between different kinds and magnitudes of impact. Thus the EIA Trial was a successful process. This point, though simple, is not to be overlooked, and is based on the evidence our review produced.

First, the EIA Final report[21] clearly outlines the various types of impact produced by participating Australian universities. By highlighting these against socio-economic objectives (SEO codes), it is clear to industry and the public where research has had the most impact, and the reasons for this. Indeed the ATN and Go8 media release of 28

[20] Australian Technology Network of Universities and Group of Eight (2012) *Excellence in Innovation Research Impacting Our Nation's Future – Assessing the Benefits*, Australian Technology Network of Universities.

[21] Australian Technology Network of Universities and Group of Eight (2012) *Excellence in Innovation Research Impacting Our Nation's Future – Assessing the Benefits*, Australian Technology Network of Universities.

November 2012 highlighted the significant economic, social and environmental benefits that have arisen from research undertaken at Australian universities.

Second, the fact that 162 case studies were produced, 64 of them by the ATN universities, shows that there was a good degree of engagement with the process itself – case study authors and institutions did not fail to produce case studies or identify impact. Our case study review shows that people could identify and articulate impact, though inevitably the process will need improvement (see Section 2.2 below, in particular). We also found that not only case study authors but also the panels were able to engage with the concept and assessment of impact, and differentiate between various types of impact. Of the 64 case studies produced by the ATN universities, over 95% were assigned a score, with just over 30% rated an 'A' or 'B', and just under half receiving a rating of 'C'. Our analysis shows there was differentiation in the scores (Figure 2.1), and although the distribution is not normal, it is certainly skewed. This indicates that panels were able to differentiate between 'good' and 'poor' impact, and validate those findings through a collaborative moderation process.

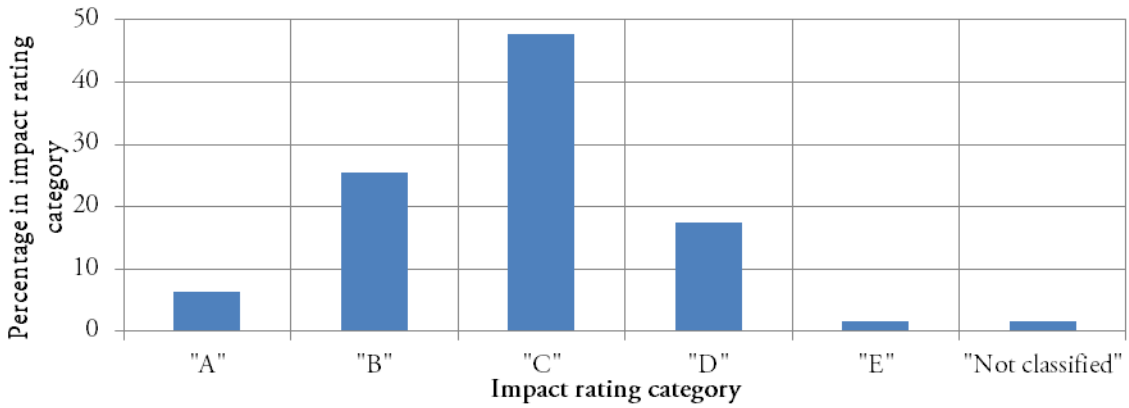

Figure 2.1: Overall distribution of case study impact category ratings for the ATN universities (excluding the 'unrated')

In addition to the ability to differentiate between various types of impact, for the six case studies double scored across the panels (three seen by the Economic Development Panel, two seen by the Society Panel and two seen by the Environment Panel)[22] there was broad agreement on scores across the panels, showing evidence of good inter-panel reliability (see Annex C):

- Two case studies received identical ratings from both sub-panels.
- Three received ratings that were within one rating point of each other.

[22] There were seven case studies that were seen by two sub-panels, but one only received one score, leaving six that actually received two scores from different panels.

- One obtained ratings that differed by two rating points.[23]

Intra-panel reliability was evident when we explored the levels of consensus within panels and the correlation between the individual panel member draft ratings and the final panel ratings for each case study. The demonstration of both intra-panel and inter-panel reliability supports the assertion that the process of the EIA Trial was relatively successful.

Finally, our analysis of the responses to the survey of institutions and case study authors showed that individuals participating in the trial were broadly positive about the process and the experience. More than three-quarters (76%, or 19/24) of the case study authors thought that the format for capturing impact was generally appropriate.[24] The case study authors identified several factors which they thought had led to the positive outcomes of the trial, including:

- appreciating the value of research at a university level and in the wider community (seven authors)
- understanding the impact and effect their research has had (12 authors)
- being able to adapt the material for other purposes (six authors)
- appreciating the need to collect ongoing evidence (four authors)
- consolidating networks through the reestablishment of contacts with clients and end users (two authors).

It is also notable that the challenges identified by institutions and case study authors were mostly linked to constructive improvements which could be made in future assessment processes. In addition to the case study authors, all five ATN institutions responded that they saw clear benefits from the trial, in particular the ability to carry the knowledge about impact forward.

Despite these successes, the process could be improved. The subsequent headlines of this report will highlight in detail various issues and ideas for improving the exercise that have emerged from our analysis.

2.2 The quality of case studies could be improved

For the assessment of case studies to be fair it is necessary for the variability in their overall scores to be based on the actual impact of the research, rather than the way in which the case study draft has been constructed or articulated. Assessed differences between case studies must be reflected in the relative impacts and any artefacts of the drafting process must be minimised. With this in mind, and based on the critical review of the guidance,

[23] One case study in the Environment Panel received a rating 'between C and D' from sub-panel 1 and a rating of 'D' from sub-panel 2; for the analysis presented here, the rating assigned by sub-panel 1 was taken to be a 'C'.

[24] In reporting the results of the case study author survey, it should be noted that the response rate was 38% (24/64), so there may be a self-selection bias in respondents. However, it could be argued that those who were both positive and negative about the EIA Trial would have equally weighted by differential reasons for responding.

the survey of institutions and case study authors, and the analysis of panel scoring, we believe there is an opportunity to improve the quality of the case studies and by so doing reduce the chance that assessed differences are artificial. To do this we propose three areas for improvement:

- to make minor changes to the guidance documentation for authors and panel members
- to adjust the case study template
- to provide material and training on how to write good impact case studies.

Each of these is explored further below.

2.2.1 Making marginal changes to the guidance documentation and its implementation

The guidance material provided to authors and panels was very concise and clear. Indeed one of its strengths was that it was considerably shorter than the REF guidance. This was confirmed in our institutional survey in which two institutions thought the guidance was 'adequate', one thought it was 'very satisfactory', one stated it was 'very good' and another that 'it was as good as could be expected'. However, when asked, all surveyed institutions identified a number of areas in which the guidance could be improved for future use. Our review and comparative assessment also identified areas for improvement. These included:

- *Timelines should be extended for completing the case studies.* When asked how the guidelines could be improved, only one institution specifically stated that the notice period should be significantly extended. However, elsewhere in the survey, four institutions noted that the notice period for the trial was too short (although one also noted that now the exercise has been undertaken once, subsequent rounds will require less preparation time).
- *The Society Panel should be divided into sub-categories*. One institution stated that the society impact area should be divided into sub-categories. This was also commented on by two other institutions elsewhere in the survey. Two institutions provided examples of how the impact area could be split more usefully, for example dividing it into health, education, law and politics and cultural understanding.
- *Greater clarity around the importance of the quality of the research underpinning the impact in the assessment process*. As noted in the critical review of the guidance in Annex A, the EIA guidelines state that impacts should be underpinned by quality research, but fails to provide a quality threshold for research. When asked how the guidance could be improved, one institution stated that this point required clarification and another institution commented that a 'suite of metrics' or alternative consistent ways to represent the quality of research would be useful elsewhere in the survey.
- *Clarity around the use of external links*. One institution noted that there should be greater clarity around the use of external links in the case study, such as embedded hyperlinks to end-user websites. It is not clear whether panels did or will be following up on these links or whether the case study author should provide contextual information summarising the content of the link. One institution

stated that the fact that external content could be provided 'meant that it was difficult to determine exactly how best to showcase the impact'.

- *Clarity around the types of impact required and associated examples*. One institution suggested that the guidance could be improved by providing more case studies from the trial that highlight the type of non-academic impact it is necessary to document in the impact case studies. A further two institutions stated that they would appreciate best-practice examples of impact.
- *Better training and guidance for panel members to ensure consistency across the assessment panel*. One institution noted that almost all of their case studies were awarded the same mark 'even though they varied considerably in quality and impact reach'. Therefore, they suggest there should be training and guidance to ensure greater consistency on how the assessment criteria should be applied across all case studies and disciplines.

2.2.2 Making adjustments to the case study template

Two of five surveyed institutions specifically noted that reordering the impact case study template would significantly improve the overall quality of the case study and another stated that the chance for repetition exists in sections. One institution suggested that the outline of the underpinning research should precede the outline of impacts in order to aid the case study author in providing a logical pathway to impact. We also drew this observation from our critical review of the guidance and it was noted by three of the 24 case study authors who responded to our survey. Two authors noted the overlap of information provided in different sections. The use of the context and summary sections appear to require further guidance, given that they were not well written in the case studies we reviewed. Finally, two institutional survey responses revealed confusion around completing the sections in regard to the use of external links and the required substance for each section.

2.2.3 Providing material and training on how to write a good impact case study

As noted above in relation to guidance improvement, it may be beneficial to provide best-practice examples of impact case studies or practical advice, neither of which was provided for the EIA Trial. This was further confirmed in our review of a sample of ten case studies where we identified a number of 'errors' in the application of the guidance, which may have been overcome through training and/or the provision of example case studies. These include case studies claiming future (not realised) impacts, conflation between academic and wider societal impacts, insufficient justification for the impacts claimed, use of the context section, and adequately demonstrating the quality of research. In three of the ten cases the authors claimed future, potential impacts and it appeared to us that this was not always picked up by the assessment panels. For example, one case study listed the expected value of economic gain from their research and this was accepted by the scoring panel and cited as evidence of impact. In two other cases, academic and societal impacts were conflated; for example, the authors claimed impact from the presentation of research findings and/or the publishing of journal articles or books. In five out of ten cases, our review team felt that the case study lacked evidence to confirm and justify the stated impacts. In all but three cases, the context section was not correctly used; and only four case studies demonstrated the quality of their research. This was achieved through

descriptions, download figures, impact factors, Google and Scopus citations. We also note that at times these deviations from the guidance were seemingly overlooked by the panels.

Of the ten case studies we reviewed, eight had scores (two were unclassified) and of those eight we disagreed with the panel assessment in one instance. The basis of this disagreement was that although we agreed with the panel that it was unclear whether there has been a big impact on the final beneficiary of the research, it would appear that the panel has not taken into consideration the influence of the long-standing research that led to the Australian Government implementing a change in policy. Finally, the idea of providing best-practice examples of impact and case studies arose spontaneously from three of the five institutions surveyed and the need for guidance and support was outlined by seven of the 24 authors.

Although it appears to be relatively straightforward to implement the changes to the guidance and template as suggested above, it may be beneficial to do so through a consultative process. Less straightforward will be the provision of exemplar case studies and possible training courses or 'how to' guides. There may be potential privacy issues with regard to the publication of the Trial case studies and if the guidance and templates are changed as suggested they will become redundant. However, the material from these case studies could be re-worked into a new template allowing potential participants of a future assessment to have model case studies to work from.

2.3 Consider including the contribution of research as an assessment criterion (in addition to reach and significance)

There are two main criteria in the EIA guidelines for assessing the nature of research impact: reach and significance, where reach is the spread or breadth of influence or effect on the relevant constituencies, and significance is the intensity of the influence or effect – which is consistent with the UK REF impact assessment criteria.[25] However, neither the EIA nor REF guidance align reach and significance with contribution, either for the contribution of the research to the impact or the contribution of various research groups to the overall research project (and hence the impact). The consideration of contribution is important for a number of reasons. First, although the impacts claimed in a case study could be large, it may be that the contribution of the claiming researcher or research group to impact is relatively small. This should be taken into account when assessing impact, and could be addressed through the implementation of a new scoring rubric, which is outlined below (as well as reordering the case study template in order to enable a logical 'research to impact' narrative and producing a 'how to guide' as discussed in section 2.2 above). This point is particularly important in attempting to assess the link between quality research and impact, as the EIA hopes to do. Without a means of understanding how research, regardless of quality, is contributing to the impact, it will be difficult to undertake further causal analysis of the kind that is sought.

[25] The REF developed the dual concept of 'reach' and 'significance', although other evaluation frameworks such as the Research Assessment Exercise and the Standard Evaluation Protocol use the term 'significance' in relation to research excellence and impact.

Second, contribution should be acknowledged in order to address the issue of double counting, whereby several collaborating institutions claim the same impact. However, double counting only becomes problematic if the exercise is performed for reasons other than advocacy and accountability – such as a means to allocate funds. In this case, it is essential to establish a way of determining the contribution to research of each institution, in order to align contribution with the reach and significance of the impact.

There are several pieces of evidence from across the evaluation workstreams that support this observation. First, the critical review of the EIA guidance documentation found that there is no explicit information on how case study authors and panel members should address any contribution issues, although both sets of guidance highlight the importance of addressing contribution. Both the EIA and the REF guidelines state that 'where the submitted unit's research was part of a wider body of research that contributed to the impact (for example, where there has been research collaboration with other institutions), the case study should specify the particular contribution of the submitted unit's research and acknowledge other key research contributions'.[26] Given that the guidance for the REF and the EIA provide additional detail about the reach of assessment criteria – 'the spread or breadth of influence or effect on the relevant constituencies' – and significance - 'the intensity of the influence or effect' – it seems logical to provide similar guidance about the way in which the specific contribution of research should be addressed.

Despite the EIA guidance stating that researchers should address contribution issues in their case study, not a single case study reviewed for this evaluation mentioned the contribution of their research to the impact, nor that of any collaborators. Moreover, neither the institutional nor case study author survey responses raised contribution as a difficult issue in the pilot. This implies that either contribution was not identified as an important concern (suggesting that the guidance should be amended to emphasise its importance) or that case study authors did not know how it should be addressed.

Second, there is no way to account for contribution of the research in the scoring guidance given to panels and no evidence that this was accounted for in any way by the panel members during their assessments. In fact, our analysis of the individual panel scoring sheets found that there were significant correlations between the way reviewers rated the descriptions of how the research underpinned the impact, the nature and extent of the impact, and the validation of the impact, but we do not know whether these descriptions were evidenced within the case study itself. Our review of the case studies showed that in five out of ten the descriptions of the impact were sufficient but in our view the evidence was lacking. In addition, we do not know how this was, or could have been, accounted for by the reviewers.

Our third supporting observation is the fact that the EIA guidance does not set a quality threshold for underpinning research as the REF does. This makes it difficult to achieve the aims of the EIA in linking the quality of research to impact, and without the basis on

[26] Australian Technology Network of Universities and Group of Eight (2012) *Excellence in Innovation Research Impacting Our Nation's Future – Assessing the Benefits*, Australian Technology Network of Universities, Appendix 5, p. 17; Research Excellence Framework (2011) *Assessment Framework and Guidance on Submissions*, REF 02.2011, p. 53.

which to understand how the research contributed to the impact in the first place, the quality of the research becomes a secondary issue. The relationship between the quality of the research and its impact is complicated, but clearly it is important that poor quality research is not rewarded, especially as it is not part of the remit of universities to carry out such research. Determining how to assess research quality is not straightforward, though, and metrics such as peer review funding or journal publications may not be sufficient. There are a number of anecdotal examples of research that has not been judged to be of sufficient quality for funding by peer review panels – for example the observation that Helicobacter Pylori was present with chronic gastritis and gastric ulcers[27] and early work in the development of *in vitro* fertilisation (IVF).[28] Likewise there have been examples of research that has been initially accepted for journal publication, but turns out to be poor research and although it was retracted nevertheless had a very large impact, as in the case of the association between the MMR (measles, mumps and rubella) vaccination and autism.[29]

In light of these observations it may be beneficial to amend the guidance for case study authors and panel members in order to address explicitly how contribution of the research to the impact is accounted for. To this end, we suggest that a scoring matrix such as the one outlined below is used. The rubric works by first assessing the significance and reach of the case study to identify the 'space' for the impact to be scored (Figure 2.2). It then considers the contribution of the research to that impact (Figure 2.3) and assigns a score so that the final score encompasses all three 'layers'. In practice it is important to note that this is not suggesting a linear model so it will be important to ensure that there is a threshold level of impact before assessing contribution (to avoid inflating an assessment of an impact that had low reach and significance but high contribution). The example shown in figures 2.2 and 2.3 demonstrates that including the low contribution of the institution's research to the impact, the score decreased by two 'grades', from an 'A' to a 'C'.

[27] Warren, J.R. and Marshall, B. (1983) 'Unidentified curved bacilli on gastric epithelium in active chronic gastritis', *Lancet* 321(8336), pp. 1273–5.

[28] Horrobin, D.F. (1990) 'The philosophical basis of peer review and the suppression of innovation', *Journal of the American Medical Association* 263, pp. 1438–41; Horrobin, D.F. (1996) 'Peer review of grant applications: a harbinger for mediocrity in clinical research?', *Lancet* 348(9037), pp. 1293–5.

[29] See Wakefield, A. et al. (1998) 'Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children', *Lancet* 351 (9103), pp. 637–41; followed by Editors of the Lancet (2010) 'Retraction – ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children', *Lancet* 375 (9713), p. 445.

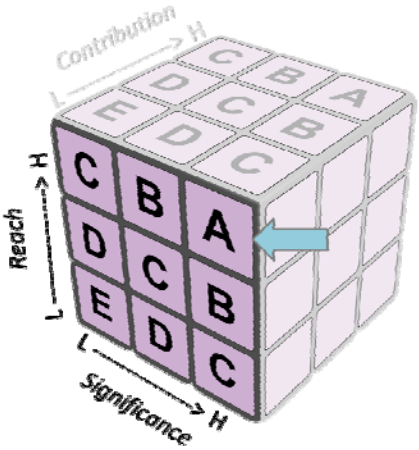

Figure 2.2: Impact score is composed of 'high' significance and 'high' reach

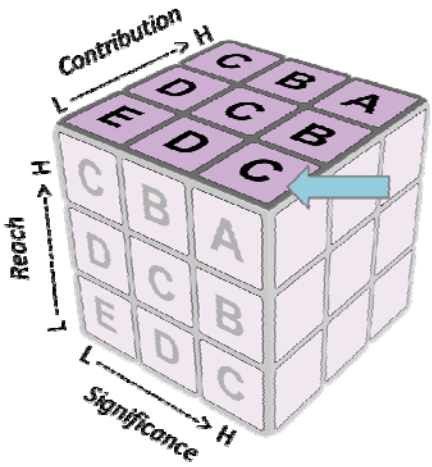

Figure 2.3: Contribution of research to impact is considered

It is also worth reflecting on the discussion in the introduction about identifying the 'contribution' by asking the counterfactual question: 'what would have happened if the contributor's research did not occur?' Research is often incremental, collaborative and iterative, so isolating the contribution of a particular piece of research to a given outcome is challenging and relies on some form of judgement, hence one of the key reasons why expert panels are best placed to make a judgement on research impact. The purpose of including contribution as an explicit assessment criterion is to encourage case study authors to attempt to demonstrate the counterfactual argument, and by doing so help panel members reach a fair assessment.

2.4 There is a need, and an opportunity, to create and embed a culture of and expectation for impact with Australian universities and wider society

Interest in the assessment of research impact is increasing internationally.[30] This is linked to a growing accountability agenda, driven by the demand for good governance and management, fiscal austerity in a number of countries and a deepening understanding of the wider role of higher education institutions in society. Therefore, there is a need to demonstrate accountability for the investment of public funds in research. This has been complemented by a shift to more comprehensive research evaluations, which encompass measures to assess wide societal impacts, as well as the standard measures often associated with academic impact such as the number and quality of publications for example. The EIA Trial is a good example of such a comprehensive evaluation. While these international trends are relatively new, it is also worth noting that Australia has been at the forefront of developing national research evaluation and impact assessment systems over the past ten years. Although the Research Quality Framework (RQF) was subsequently superseded by the ERA, when it was originally proposed its aim was to assess the quality and impact of research in Australia. Indeed much of the technical work that was undertaken for the RQF was used in the development of the impact assessment element of REF.[31] The bibliometric analysis performed for the ERA is probably the largest exercise of its kind worldwide.[32]

All of this suggests that the concept of accountability that goes beyond the academic system is relatively well founded within the Australian higher education system. Nevertheless, the EIA Trial served to highlight the need for institutions to create systematic mechanisms to capture impact. Indeed, three of the five surveyed institutions stated that the need to develop mechanisms to capture impact regularly was a key insight from the Trial, and one institution has already developed such a system. Moreover, eight of 24 case study authors stated that they valued the opportunity to think about their impact and seven appreciated the chance to demonstrate the value of their research.

Looking to the future, it will be important for the higher education sector to continue to create and embed a culture and expectation of impact – not only within the university itself, but also among the tax paying public and their representatives. Exercises such as the Trial are an excellent first step in this regard. Institutions should be encouraged to develop internal impact strategies that focus on not only operational aspects of impact assessment – including for example institutional approaches to the information management of impact – but also how to maximise impact as and when it occurs. At the same time it is important to acknowledge that achieving and maximising impact is sometimes out of the researchers' and institutions' sphere of influences and there is often an element of opportunity moving along the research-to-impact continuum. That said, institutional strategies should be supported by wider policy initiatives such as the performance-based allocation of funding in relation to wider impacts. It is particularly interesting to observe that one surveyed

[30] Guthrie, S. et al. (forthcoming) *Measuring Research: A Guide to Research Evaluation Frameworks and Tools*, Santa Monica, CA: RAND Corporation.

[31] Grant, J. et al. (2010) *Capturing Research Impacts: A Review of International Practice*, Santa Monica, CA: RAND Corporation.

[32] Australian Research Council (2012) *Excellence in Research for Australia (ERA) 2012 National Report*.

institution noted that the burden of the EIA Trial would be justified if funding was a potential outcome.

2.5 Steps should be taken to understand the benefits and (unintended) consequences of future assessments

The final issue related to improving the EIA Trial in its current form is the need to understand the (unintended) consequences of additional future assessments, alongside the benefits. Benefits might include the continued awareness and understanding of what kinds of impacts arise from the ATN universities and the research they conduct. These benefits will likely accrue for the researchers within the universities, institutions and industry, and among policymakers, research stakeholders and the wider public. Understanding and highlighting the impacts from the research produced by Australian universities can be used to serve the purposes of any of the four drivers: for *advocacy,* to show *accountability*, as a means of *allocating* funding, and for *analysis* of what factors lead to and contribute to impact. It can also serve the purpose of supporting wider societal and policy agendas, and generally improving the awareness of why and how research serves the needs of our societies more broadly and in what areas.

However, future assessments of impact will face additional challenges, in particular if the exercise is scaled up as discussed in the next chapter. The idea that the impact of research both increases and decreases over time provides a challenge if the impact assessment is repeated over a number of years. Should the exercise be repeated, it needs to be decided whether the same case study will be eligible for submission several times providing that the impacts have changed since the last submission. If repeat submissions are feasible then a decision needs to be made on whether the impact assessed is the total or marginal impact between cycles. There will need to be a strong rationale if case studies cannot be re-submitted as there will be consequences to effectively 'side-lining' entire areas of impact.

Second, future assessment exercises could lead to 'gaming', whereby researchers manage impact, for example, by generating media interests to fit impact assessment timelines and windows. This would undermine the very purpose of an impact exercise, which should be to promote positive engagement with impact and improve understanding of how to maximise it, rather than to manage it proactively to fit into a bureaucratic exercise. It is important to recognise that researchers may reasonably object to being asked to divide their research into arbitrary impact time periods. This may lead them to delay or accelerate research outputs to fit into research assessment cycles, potentially threatening the underlying research and its value for future academic research and teaching.

Moreover, it should also be acknowledged that it is not essential for every research project to have an impact. Indeed, basic research has a justifiable place in society and may not result in wider societal impacts of the kind assessed in this trial or within the timeframe of assessment.[33] Nevertheless, our experience suggests that an impact assessment exercise such as the EIA Trial can have positive spillover effects by encouraging researchers to consider

[33] Salter, A.J. and Martin, B.R. (2001) 'The economic benefits of publicly funded basic research: a critical review', *Research Policy*, 30(3), pp. 509–32.

more proactively how their research may benefit wider society, but without distracting from the core purpose of conducting research. The survey responses clearly demonstrate this point given that many authors and institutions reported their appreciation for identifying and understanding impact increased as a result of the Trial. This 'culture of impact', as previously discussed, needs to be positively managed so that gaming, such as that seen in previous manifestations of the Research Assessment Exercise and ERA in relation to publication outputs,[34] does not occur with the impact exercise.

Finally, there are consequences and benefits for wider stakeholder engagement, which should be considered in relation to the time and resources stakeholders contribute to the assessment panels, and the consequences and resources required to engage them in identifying impact. Given that any potential funding will not flow to end users there is very little direct benefit of their participation in assessing research impact. However, the involvement of stakeholders in providing evidence is essential, and may result in positive benefits for the stakeholder and the researcher in increasing impact awareness; it will also act as an additional demand on time and resource. Section 3.2 outlines how these kinds of resources may only be justifiable if the exercise is used to allocate funding.

In order to address these issues, a decision should be taken soon about the future of the exercise and clear and transparent guidance should be issued that takes into account the concerns raised in this report. It may be appropriate to establish a 'blue ribbon panel' to think through the nature of the (unintended) consequences and be explicit and transparent about them with the research and wider communities that will be affected.

[34] Many authors and commentators have written on this subject. See for example, Roberts, G. (2003) *Review of Research Assessment*, London: Higher Education Funding Council for England; Hicks, D. (2011) 'Performance-based university research funding systems', *Research Policy* 41: 251–61; and Macilwain, C. (2010) 'World view: wild goose chase', *Nature* 463, p. 291.

CHAPTER 3

Issues and ideas on how could the EIA Trial be scaled up as a companion piece to ERA

In this chapter we set out the three headline observations that would need to be considered if EIA was to be scaled up as a companion piece for ERA. As noted, along with the observations described in the preceding chapter focusing on improving EIA if it were to be repeated, these are not mutually exclusive. All eight should provide the basis for continued discussions and consultations about how to develop the solid foundation for impact assessment.

3.1 The EIA impact assessment process will need to be reconfigure if scaled up

If the EIA Trial is scaled up and used as a complement to the ERA, it will need to be reconfigured in order to take into account three distinct, but complementary issues: the primary purpose of the assessment; the appropriate unit of assessment for a larger scale; and limitations on the number of case studies that can feasibly be reviewed by the panels. The reasons underpinning each of these issues and the relevant evidence across all workstreams are discussed in turn.

3.1.1 Primary purpose of the assessment

If scaled up, the primary purpose of the EIA assessment may change from one which was about understanding 'the dividends' of research and innovation in Australian universities, to one which is used as a complement to the ERA and, potentially, a means of allocating funding. A change in the strategic intent of the EIA may require changes in the detailed process within the established EIA framework, including how the unit of analysis should be selected, and how attribution and contribution are determined and rewarded.

If the EIA is linked to funding decisions in the future, the guidance should be explicit in how performance is related to funding at the outset and the approach should provide a system for the unambiguous rating and comparison of submissions. Both a case study author and an institutional survey response noted that the burden required from the EIA would be justifiable if the exercise was linked to funding. Therefore, even the nature of how people engage with the exercise may be affected by the purpose and aims.

3.1.2 The unit of analysis

There are two issues to consider. The first is the level of (dis)aggregation for assessment units and the second is the challenge of 'multi-impactful' research. The EIA Trial used

four-digit SEO codes as its main organisational unit of analysis. As noted in the introduction, this is a unique characteristic of the EIA Trial when compared to the UK REF. It also complements the ERA approach of using fields of research (FOR) codes rather than departmental driven disciplinary units within universities (as with the UK REF).

In EIA the four-digit SEO codes were grouped into four main panels – on defence, economic development, environment and society. These are very broad and this could pose problems for a scaled up trial, unless some form of sub-panel structure is introduced, because each main panel will need to be large enough, and diverse enough in its composition, to assess all case studies submitted to it. If we assume that case study submissions are determined by the total number of full-time equivalent (FTE) staff submitted to ERA 2012 (c. 42,000),[35] and use a similar formula to that used in the UK REF (one case study is required for every ten submissions), then the indicative number of case studies submitted could be close to 4,200. This is not dissimilar to the 52,400 staff submitted to the UK 2008 Research Assessment Exercise,[36] suggesting different strategies in submitting research staff with the 4,200 being a likely overestimate. An alternative approach for estimating the number of case studies would be to pro rate the likely number of UK case studies (about 5,000) on population size and the number of higher education institutions. This generates estimates of 1,458 and 1,250 case studies respectively, suggesting a likely workload of no more than 1,500 case studies for planning purposes.[37]

Either way this suggests that the current structure of four main panels (seven including sub-panels) will need to be reconfigured, either by having more sector based panels with narrower remits or by having a hierarchical system of multiple sub-panels associated with four-digit SEO codes (or a greater number of aggregated combinations than used in the EIA Trial). However, this will then raise the key issue of inter-panel reliability, which will need to be assured through guidance, briefing and training of panel members as discussed in the next section.

In raising this issue we fully recognise that the use of SEO codes to make impact central to the assessment was viewed by many as one of the strengths of the EIA Trial. Indeed we are not arguing that SEO codes should not be used as the organisational principle, rather that thought and discussion is needed on the number of SEO code defined panels and their interrelationship with one another. In our survey, two of five institutions stated that SEO codes are appropriate units of analysis as opposed to research disciplines, and another commented that the analysis groupings were acceptable. However, three of the five surveyed institutions explicitly stated that the four primary sectoral divisions under which case studies must be submitted are unhelpful. More specifically, all three institutions felt

[35] Australian Research Council (2012) *Excellence in Research for Australia (ERA) 2012 National Report*. It should be noted that the ERA focuses on the current researcher population, while EIA provides a retrospective view of that population in the past. The implications of this methodologically if the two are meant to be linked will need to be thought through. Indeed, the issue of where researchers have currently been located as opposed to where they are now is, perhaps, an unnecessarily complicating factor in the UK REF.

[36] Research Assessment Exercise (2008) 'RAE2008 confirms UK's dominant position in international research'.

[37] The ratio of Australian population to UK population is 21.5m/61.9m = 0.35 x 5000 = 1458; the ratio of the number Australian higher education institutes to UK higher education institutes is 41/165 = 0.25 x 5000 = 1250.

that the Society division was far too broad, and one felt that that the Defence division was too narrow (as did one case study author). Two institutions also suggested that there should be guidance regarding how case studies should be approached when they apply to several of the four primary categories.

The second issue with using SEO codes is that they may exacerbate the risk of under-estimating and under-representing the diversity of impacts that could originate from one area of research. One way to think about this is to ask whether research is more likely to be multi-*disciplinary* in nature, or multi-*impactful*. Consider Figure 3.1 as representative of the 'extremes' of the different ways an impact pathway might unfold. The blue triangle is one extreme where research starts from multiple disciplinary origins but only has a narrow set of impacts at the end. The other, the red triangle, is that research originates from one type of discipline, but goes on to have many different kinds of impacts.

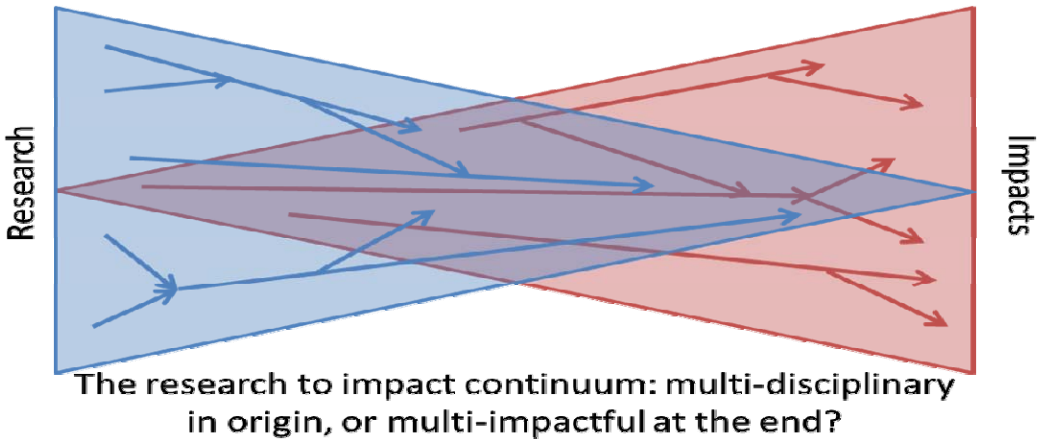

Figure 3.1 : The theoretical research-to-impact continuum

Obviously neither situation is representative of reality and most types of research probably have some of each situation, but again it is essential to consider whether it is more likely that research will have multiple inputs or multiple impacts. The organisational unit of analysis should arguably be based on the area that is less likely to be multi-faceted. In our experience, it is more likely that research will have multiple impacts, which presents an argument for either allowing multiple impact submissions arising from a signal body of research (as the EIA does) or organising the exercise from the starting point of research disciplines or areas. This latter point enables all impacts to be represented through the case study and the authors do not need to highlight only particular types of impacts. In other words, in the EIA Trial while the intention of the SEO codes may encourage researchers and institutions to think about the various kinds of impacts that might have arisen from research, it may have the unintended effect of broadening out the research base that fed into the impacts, but narrowing down the types of impact discussed. The resolution of this point is linked to other areas highlighted in this report, such as the ultimate aim of the EIA, and as noted earlier we raise this as a point for further discussion and debate as the conversation of developing the EIA unfolds.

3.1.3 Limit to the number of case studies which can feasibly be reviewed

Finally, if the EIA Trial is scaled up, it will be necessary to reconsider the way in which panels assess case studies. Simply put, there could be as many as 1,500 submitted case

studies, requiring four main panels to assess close to 375 case studies each (assuming an equal distribution). Of particular note is that one of the Society sub-panels could not assign ratings to seven case studies because it ran out of time.[38] Many more sub-panels will certainly be required in a scaled up exercise and running out of time will not be an option.

The main concern arising from this issue is inter- and intra-panel reliability. Although our analysis suggests that this may not be particularly problematic, the issue should not be disregarded. As discussed in Annex C, inter-panel reliability for the six case studies reviewed by two sub-panels was good. In addition, intra-panel reliability was also good as the majority of all panels had levels of consensus over 75% between individual panel member ratings and the overall panel score. Moreover, no panel was significantly more favourable or unfavourable in their ratings when an inter-panel comparison was performed. However, if SEO codes cut across main panel areas (e.g. environment and economic development), there will be questions as to which panels will review which case studies and inter-panel reliability across sectors could be an issue.

Though the burden on panel members was not mentioned in the survey responses[39] and we did not survey the panel members themselves, we note from our review of the guidance and related documents that the EIA mentions the considerable (largely voluntary) time put into the panel assessments by members. Again, in a scaled up exercise this amount of resource will be considerable and to this end the institutions we surveyed thought it was unreasonable that panels would have time to review all of the information in the case studies, such as external evidence links. If there are concerns about such fundamental underpinning information such as evidence for claimed impacts not being reviewed, careful thought needs to go into how the panels would work in a scaled up exercise and what is feasible, and reasonable, to expect of them.

We therefore suggest that the way to implement all three of these issues is to consider a working group or blue ribbon panel, which would consider the host of issues we have raised here, as well as many others, and develop a clear and structured plan for how the new exercise will be run, and more crucially what its purpose will be and how it will translate into the organisational unit of analyses and assessment processes. There will need to be wide consultation on these plans and guidance so there is buy-in across the sector, within the higher education and research policy community, and crucially with end users and other stakeholders given their buy-in will be key in scaling up the exercise.

[38] Although it should be noted this was largely to discussions on interpreting the guidelines reinforcing the earlier observation that better briefing and training of panels would have alleviated at least some of the time issues.

[39] Although one institutional survey respondent did note that it is unreasonable to assume panel members will have the time to access external links.

3.2 Any future EIA impact assessment will need to ensure that the direct and indirect benefit of assessing impact in Australian universities outweighs the transaction costs of the assessment

It is essential for any scaled up impact assessment system to provide value for money, so the costs associated with the identification, articulation and evaluation of impact case studies should be significantly less than the benefits that can arise from such a systematic assessment. Transaction costs of around 5% to 10% are typically accepted as the norm for research funding allocation evaluation exercises.[40] Therefore, the allocation of AU$100m should cost between $AU5m and $AU10m to administer. When undertaking such a cost–benefit analysis it is important to take into account both the direct and indirect costs as well as benefits:

- Direct costs include the amount of time case study authors spend producing each case study.
- Indirect costs include interactions with stakeholders outside the academic system who are able to provide evidence to support impact case studies.
- Direct benefits arise from any funding allocated based on the assessment of impact (a good example of this is the UK REF where an estimated £220m a year of quality-related funding will be allocated to higher education institutes based on impact).
- Indirect benefits include, for example, raising the profile of a research institution as identified above.

Table 3.1 presents a crude estimate of the direct transaction costs associated with the EIA Trial. We have identified three cost areas: those associated with the universities, with the panel and with running the trial. We have worked with considerable caution in arriving at these estimates and where possible have used supporting data from the workstreams. Our best estimate of the direct cost for each case study is six days. It is worth noting that these estimates are similar to those derived in the evaluation of the REF pilots.[41] For example, it was estimated that it took 3.2 days (25.6 hours) to prepare each case study, which is very similar to our reported median of 26 hours, and 2.1 days (16.8 hours) of 'coordination effort per output'[42] (again close to the 12 hours estimated in the pilot study). Unfortunately we do not have any benchmarks to compare the effort taken by the panels in reviewing the case studies, but note that the case study template used in the EIA Trial was longer than that used in the REF pilots. To monetise this cost we have assumed a day rate of AU$1,000,[43] so each case study costs around AU$6,000. Given some of the

[40] Sastry, T. and Bekhradnia, B. (2006) *Using Metrics to Allocate Research Funds: A Short Evaluation of Alternatives to the Research Assessment Exercise*, Higher Education Policy Institute.

[41] Technopolis (2010) *REF Research Impact Pilot Exercise Lessons-Learned Project: Feedback on Pilot Submissions: Final Report*, Technopolis, November 2010.

[42] The term output is used to cover both impact case studies and impact templates (which are used in REF to providing an overarching strategic statement of how impact is assured for a given unit of assessment).

[43] This estimate is crudely derived by assuming an annual average salary of $AU100,000 taking into account overheads at 100% and dividing this by 200 working days.

underlying assumptions made in this estimate we suggest that the direct cost of each case study is between AU$5,000 and AU$10,000.

Table 3.1: Best estimate of direct costs per case study

Cost item	Source	Estimated (hours/case study)
University – administration	Estimated, but informed by institutional survey (Annex D)	12 hours
University – researcher	Case study author survey (Annex D)	26 hours
Panel – case study review	Analysis of scoring sheet data (Annex C)	2 hours
Panel – meetings	Estimated	4 hours
ATN	Estimated	4 hours
Total		48 hours per case study (6 days)

If the EIA impact assessment will serve to aid the allocation of funding alongside the ERA in the future, it is possible to project the most logical level of funding allocation that should result from the impact assessment – given typical transaction costs of between 5% and 10%. More specifically, for every case study submitted for assessment some AU$100,000 would need to be available for allocation to universities. Therefore, if 1,500 case studies are submitted, around AU$150m would need to be available over the assessment cycle.[44] By comparison around AU$68m was allocated through the ERA influence part of the Sustainable Research Excellence in Universities (SRE) initiative in 2013.[45] This analysis further reinforces the early observation that the 'rule' for guiding the number of case studies in a submission will need careful consideration.

3.3 There will be a need to engage early with other institutions that did not participate in the EIA Trial in a scaled up impact assessment

The final headline of our review is that if the EIA impact assessment exercise is repeated, and in particular if it is scaled up to include institutions that did not participate in the Trial, a concerted effort to engage with institutions that did not participate in the Trial will be essential and it may be necessary to run a sector-wide trial. The reason for this is straightforward: the institutions that did not participate in this initial Trial will arguably be at a comparative disadvantage and, more generally, there will be a need for broader buy-in within the university and stakeholder communities regarding the importance of impact assessment. The EIA report and this review are important first steps in this direction, but more consultation and engagement will be needed.

[44] That is if the assessment occurs every two years then the annual allocation would need to be AU$75m.

[45] Australian Government (2012) *The Process for Determining Sustainable Research Excellence (SRE) Scheme Grant Amounts – 2012 Allocations*, Department of Industry, Innovation, Science, Research and Tertiary Education.

As mentioned above, the survey responses from institutions and 12 case study authors indicated that a major reward of the trial was the opportunity to capture and document the impact of their research. Four case study authors highlighted the importance of collecting ongoing evidence of impact as a benefit and eight (32%) case study authors noted that a reward for participating in the Trial was in thinking about and engaging in the idea of how their research translates into impact, understanding the impact of their research, and being able to evaluate and articulate it. Authors stated that the assessment had improved their ability to articulate impact, to think about how research is translated and where this process can be improved, and to identify and reflect on the impact of their work.

These benefits, though positive for participating case study authors and institutions, suggest that there were real gains for those who took part in the Trial, which could put them at a comparative advantage. This is a particular issue if wider engagement is not taken into account early in the process of scaling up the assessment, in particular in relation to the burden of collating data and writing up the case study. This is seen to some extent within the case studies produced through the trial. One-quarter (6/24) of case study authors spent less than one day (8 hours) on their case study and many of this group admitted to previously collecting the information for another purpose and merely fitting it to the case study template.

However, this could be counterbalanced by a comprehensive training and support programme for all universities. To ensure non-trial participants are not disadvantaged in future assessments, at least a selection of case studies and their scores should be made public. In fact, one case study author mentioned that he received encouragement from reading other case studies. This may be an important entry point to the process of engagement and an added incentive.

It will also be important to ensure there is broad participation in redesign from all institutions and everyone has the opportunity to provide input. In many ways we believe Australia is already far ahead of where the UK was at this point in the development of the REF. After the REF pilot, institutions and those participating in the pilot were still, to a certain extent, bemoaning the inclusion of impact. It is notable that no one we surveyed thought the EIA Trial was a fruitless or meaningless process. Rather, they were positive and enthusiastic about the lessons learned. The ATN and other organisations should harness this as the next stage of the EIA Trial moves forward.

3.4 Concluding comment

In conclusion, our review has highlighted that the EIA trial was in many ways a success and has served to bring the importance of understanding and disseminating the wider impacts of academic research into sharp focus. There are clear benefits to understanding and engaging widely within the academic and wider stakeholder communities to communicate these impacts. In several respects, we believe the Australian higher education community is at the forefront of global discussions, and further advanced in thinking through ways of how to engage in these efforts. However, there are still improvements which can be made and important contributions the whole community can make to the process. We hope this report serves to set in motion a series of broader national discussions about the different elements that contribute to a successful impact assessment exercise.

These include thinking about how to maximise some of the benefits such as building a culture of impact, encouraging wider stakeholder engagement with research, demonstrating the value for industry and society of research investments, but also minimising some of the unintended consequences. There is a real opportunity for Australian universities to be a leading example for how impact is and should be assessed.

REFERENCES

Bibliography and reference list

Academy of Medical Sciences (2006) *UK Evaluation Forum Medical Research: Assessing the benefits to society*, London: Academy of Medical Sciences. Available at: http://www.acmedsci.ac.uk/index.php?pid=99&puid=64 [last accessed 27 February 2013].

Australian Government (2012) *The Process for Determining Sustainable Research Excellence (SRE) Scheme Grant Amounts – 2012 Allocations*, Department of Industry, Innovation, Science, Research and Tertiary Education. Available at: http://www.innovation.gov.au/Research/ResearchBlockGrants/Documents/SREProcessCalculations2012.pdf [last accessed 15 March 2013].

Australian Research Council (2012) *Excellence in Research for Australia (ERA) 2012 National Report*. Available at http://www.arc.gov.au/pdf/era12/report_2012/ARC_ERA12_Section1_page25to36.pdf [last accessed 27 February 2013].

Australian Technology Network of Universities and Group of Eight (2012) *Excellence in Innovation Research Impacting Our Nation's Future – Assessing the Benefits*, Australian Technology Network of Universities. Available at: http://www.atn.edu.au/eia/Docs/ATN-Go8-Report-web.pdf [last accessed 12 March 2013].

Brutscher, P.-B., Wooding, S. and Grant, J. (2008) *Health Research Evaluation Frameworks: An International Comparison*, Santa Monica, CA: RAND Corporation, 2008. Available at: http://www.rand.org/pubs/technical_reports/TR629 [last accessed 27 February 2013].

Canadian Academy of Health Sciences (2009) *Making an Impact: A Preferred Framework and Indicators to Measure Returns on Investment in Health Research. Report of the Panel on the Return on Investments in Health Research*. Available at: http://www.cahs-acss.ca/making-an-impact-a-preferred-framework-and-indicators-to-measure-returns-on-investment-in-health-research-8/ [last accessed 27 February 2013].

Chataway, J. and Morgan Jones, M. (2011) 'Analyse this', *Public Service Review: European Union* 21, pp. 324–6.

Editors of the Lancet (2010) 'Retraction–ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children', *Lancet* 375(9713), p. 445.

European Science Foundation (2012a) *Evaluation in Research and Research Funding Organisations: European Practices*. Available at: http://www.esf.org/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/be_user/CEO_Unit/MO_FORA/MOFORUM_Eval_PFR__II_/Publications/mof_evaluation_final.pdf&t=1362150690&hash=457704d9e136093c52945d54deb4ee2954a3e792 [last accessed 27 February 2013].

European Science Foundation (2012b) *The Challenges of Impact Assessment: Working Group 2: Impact Assessment ESF Member Organisation Forum on Evaluation of Publicly Funded Research*. Available at: http://www.esf.org/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/be_user/CEO_Unit/MO_FORA/MOFORUM_Eval_PFR__II_/Publications/WG2_new.pdf&t=1362155294&hash=9ca06c34938cdc14b2966c170d3e0583c2c31752 [last accessed 27 February 2013].

Grant, J., Brutscher, P.-B., Guthrie, S., Butler, L. and Wooding, S. (2010) *Capturing Research Impacts: A Review of International Practice*, Santa Monica, CA: RAND Corporation. Available at: http://www.rand.org/pubs/documented_briefings/DB578 [last accessed 27 February 2013].

Grant, J. and Wooding, S. (2010) *In Search of the Holy Grail: Understanding Research Success*, RAND_OP295, Cambridge: RAND Europe.

Guthrie, S., Wamae, W., Diepeveen, S., Wooding, S. and Grant, J. (forthcoming) *Measuring Research: A Guide to Research Evaluation Frameworks and Tools*, Santa Monica, CA: RAND Corporation.

Hicks, D. (2011) 'Performance-based university research funding systems', *Research Policy* 41: 251–61.

Horrobin D.F. (1990) 'The philosophical basis of peer review and the suppression of innovation', *Journal of the American Medical Association* 263, pp. 1438–41.

Horrobin, D.F. (1996) 'Peer review of grant applications: a harbinger for mediocrity in clinical research?', *Lancet* 348(9037), pp. 1293–5.

Macilwain, C. (2010) 'World view: wild goose chase', *Nature* 463, p. 291.

Marburger III, J.H. (2005) 'Wanted: better benchmarks', editorial, *Science* 308(5725), pp. 1087.

Marjanovic, S., Hanney, S. and Wooding, S. (2009) *A Historical Reflection on Research Evaluation Studies, Their Recurrent Themes and Challenges*. Cambridge, UK: RAND Europe. Available at: http://www.rand.org/pubs/technical_reports/TR789.html [last accessed 17 March 2013].

Medical Research Council (2010) *Outputs, Outcomes and Impact of MRC Research: Analysis of MRC e-Val Data*. Available at: http://www.mrc.ac.uk/consumption/groups/public/documents/content/mrc008191.pdf [last accessed 27 February 2013].

OECD (2010) *OECD-Norway Workshop on Performance-based Funding for Public Research in Tertiary Education Institutions.* Available at:

http://www.oecd.org/document/55/0,3343,en_2649_34293_45711095_1_1_1_1,00.html [last accessed 27 February 2013].

RCUK (2006) *Report of the Research Councils UK Efficiency and Effectiveness of Peer Review Project*, Swindon: Research Councils UK. Available at: http://www.rcuk.ac.uk/documents/documents/rcukprreport.pdf [last accessed 22 February 2013].

Research Assessment Exercise (2008) 'RAE2008 confirms UK's dominant position in international research'. Available at http://www.rae.ac.uk/news/2008/results.asp [last accessed 27 February 2013].

Research Excellence Framework (2010) *Units of Assessment and Recruitment of Expert Panels*, REF 01.2010. Available at: http://www.hefce.ac.uk/research/ref/pubs/2010/01_10/ [last accessed 27 February 2013].

Research Excellence Framework (2011) *Assessment Framework and Guidance on Submissions*, REF 02.2011. Available at: http://www.hefce.ac.uk/research/ref/pubs/2011/02_11/ [last accessed 27 February 2013].

Research Excellence Framework (2012) *Panel Criteria and Working Methods*, REF 01.2012. Available at: http://www.ref.ac.uk/pubs/2012-01/ [last accessed 27 February 2013].

Roberts, G. (2003) *Review of Research Assessment*, London: Higher Education Funding Council for England.

Salter, A.J. and Martin, B.R. (2001) 'The economic benefits of publicly funded basic research: a critical review', *Research Policy*, 30(3), pp. 509–32.

Sastry, T. and Bekhradnia, B. (2006) *Using Metrics to Allocate Research Funds: A Short Evaluation of Alternatives to the Research Assessment Exercise*, Higher Education Policy Institute. Available at: http://www.hepi.ac.uk/files/23RAEandmetricsfullreport.pdf [last accessed 15 February 2013].

Slote Morris, Z., Wooding, S. and Grant, J. (2011) 'The answer is 17 years, what is the question: understanding time lags in medical research', *Journal of the Royal Society of Medicine*, 104(12), pp. 510–20.

Smith, S., Ward, V. and House, A. (2011) '"Impact" in the proposals for the UK's Research Excellence Framework: shifting the boundaries of academic autonomy', *Research Policy*, 40(10), pp. 1369–79.

Technopolis (2010) *REF Research Impact Pilot Exercise Lessons-Learned Project: Feedback on Pilot Submissions: Final Report*, Technopolis, November 2010. Available at: http://www.hefce.ac.uk/research/ref/pubs/other/re02_10/[last accessed 27 February 2013].

Torres-Reyna, O. (2013) *Getting Started in Logit and Ordered Logit Regression: Data and Statistical Services*, Princeton University. Available at: http://dss.princeton.edu/training/Logit.pdf [last accessed 27 February 2013].

Trochim, W., Kane, C., Graham, M.J. and Pincus, H. (2011) 'Evaluating translational research: a process marker model', *Clinical and Translational Science* 4(3), pp. 152–62.

Wakefield, A., Murch S., Anthony, A. et al. (1998) 'Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children', *Lancet* 351(9103), pp. 637–41.

Warren, J.R. and Marshall, B. (1983) 'Unidentified curved bacilli on gastric epithelium in active chronic gastritis', *Lancet* 321(8336), pp. 1273–5.

ANNEXES

Annex A : Review of EIA impact assessment guidelines

As discussed in the introduction to this report, the challenges of undertaking research evaluations have been broadly recognised, and many of these challenges are of particular importance to the EIA impact assessment exercise. In order to outline how these challenges may be most efficiently addressed, they have been organised into strategic and practical categories, forming criteria against which to benchmark the guidelines. Observations addressing the challenges have been offered to improve the current exercise. The guidelines have also been compared to the UK's Research Excellent Framework (REF), as well as other research impact evaluation frameworks where applicable.

A. 1. Conceptual challenges

Conceptual challenges refer to those that have a fundamental impact on how the research assessments are conceived and understood. This section will outline four challenges, which at the very least require acknowledgement in evaluation guidelines and in some cases should be addressed to significantly improve the exercise.

A.1.1 Time lags

The notion that time lags exist between the completion of research and its eventual impact has been widely acknowledged, and the attempt to measure this time lag has been the subject of various studies.[46] This is an important point to acknowledge in developing research evaluation guidelines.

The EIA impact assessment trial guidelines note this point. They state that the claimed impact should occur between 1 January 2007 and the date of submission for the Trial, and the underpinning research should have been conducted either during this time or between 1 January 1992 and 31 December 2006.[47] This is very similar to the REF guidance, which states that impacts should occur between 1 January 2008 and 31 July 2013, and the underpinning research

[46] Slote Morris, Z., Wooding, S. and Grant, J. (2011) 'The answer is 17 years, what is the question: understanding time lags in medical research', *Journal of the Royal Society of Medicine*, 104(12), pp. 510–20; Trochim, W. et al. (2011) 'Evaluating translational research: a process marker model', *Clinical and Translational Science* 4(3), pp. 152–62.

[47] Australian Technology Network of Universities and Group of Eight (2012) *Excellence in Innovation Research Impacting Our Nation's Future – Assessing the Benefits*, Australian Technology Network of Universities, Appendix 5.

should have been conducted during the period 1 January 1993 to 31 December 2013. However, it is important to note that the EIA guidance states that '[e]xceptions to this time frame may still be allowed if justified in the application'.[48] This differs from the REF guidance, which only makes an exception for the *research* window in unit of assessment (UOA) 16 (Architecture, Built Environment and Planning) in main panel C, stating:

> This time frame will be extended by five years, so that the eligibility period for research underpinning case studies in that UOA is 1 January 1988 to 31 December 2013. The main panel recognises the extended time frame is necessary, in some cases, for changes to the built environment to be delivered in practice, based on the findings of research from some areas of planning and architectural practice.[49]

With this in mind, the ability to justify an extension to the *research or impact* window in the EIA guidelines is a significant caveat, which requires attention. It is unclear from the guidance what the content and length of this justification should be. For example, it may be assumed by case study authors that outlining significant and far-reaching impacts falling outside this time window in itself justifies their inclusion. Indeed, as highlighted in Annex B, several case studies referred to impacts outside the specified impact window (including future impacts), suggesting confusion around this point. Not only this, but the very concept of separate research and impact windows may be misunderstood, and may require further clarification in the guidance.

Observations to improve the current exercise

In order to improve the impact assessment in its current form the confusion concerning the impact window should be addressed. First, it may be beneficial to include a figure such as that shown in Figure A.1, which provides a visual representation of the research and impact windows. Second, it may also be useful to provide examples of how the presentation, content and length of the impacts falling outside the specified impact window may be justified.

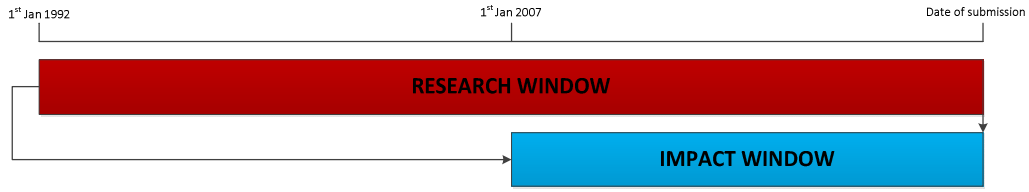

Figure A.1: Visualisation of the research and impact windows

Observations to aid scalability

The idea that the impact of research both increases and decreases over time provides a challenge if an impact assessment is repeated over a number of years. Should the exercise be repeated, it needs to be decided whether the same case study will be eligible for submission several times providing that the impacts have changed since the last submission. If repeat submissions are feasible then a decision needs to be made about whether the impact assessed is the total or marginal impact between cycles. This may also help to address the problem of current case studies referring to future impacts.

---

[48] Australian Technology Network of Universities and Group of Eight (2012) *Excellence in Innovation Research Impacting Our Nation's Future – Assessing the Benefits*, Australian Technology Network of Universities, Appendix 5, p. 6.

[49] Research Excellence Framework (2012) *Panel Criteria and Working Methods*, REF 01.2012, p. 73.

The repetition of the exercise may also cause other problems concerning the impact window. It is important to recognise that researchers may resent dividing their research and its impacts into arbitrary time periods, and this may also lead to delaying or accelerating research in order for impacts to be eligible in several 'impact cycles'. This will need to be considered if the exercise is undertaken on a larger scale.

A.1.2 Unit of analysis

The level at which impact assessments are carried out has significant implications for the conclusions and recommendations that are possible as a result of the exercise.[50] High levels of aggregation are generally thought to be most useful for assessing downstream impacts, and low levels of aggregation can cause difficulties in the attribution of outcome and impact measures.[51]

Unlike the Excellence in Research for Australia (ERA) where the primary unit of analysis is the research discipline across an institution, in the EIA trial the unit of analysis is the socio-economic objective (SEO) codes represented by the case study across the institution. Case studies were categorised using the four-digit SEO code, and from these four overarching sectors (defence, economic development, society or the environment) were defined and used as the basis for organising the panels. While it is expected that each submission will be encapsulated within one SEO classification, provision is made in the guidelines for the use of up to three SEO classifications per case study. This sectoral approach enables a high level of aggregation, given that it allows for submissions from institutions or collaborations working across disciplinary boundaries. Indeed, the EIA guidelines state that the attribution of the outcome is for the institution to determine – whether school, research centre or institute, or some combination of researchers working across institutional boundaries.[52] This is reflected in the case studies, which tend to present a body of work from an institution rather than a particular research project for which there is one principal investigator. However, it is important to note that this may be because a small number of case studies were requested from each institution, and if the exercise is repeated with more case studies from each institution, the unit of analysis may automatically disaggregate.

The sectoral approach marks a fundamental difference from the UK REF, whereby the units of analysis are divided into four main panels by subject area, of which there are 36. For example, main panel A includes clinical medicine, public health, health services and primary care, allied health professions, dentistry, nursing and pharmacy, psychology, psychiatry and neuroscience, biological sciences and agriculture, and veterinary and food science. This approach is beneficial in clearly outlining where each case study should be submitted, and the lack of a quota for each subject area per research institution allows each case study to be submitted under its most relevant category.

---

[50] European Science Foundation (2012b) *The Challenges of Impact Assessment: Working Group 2: Impact Assessment ESF Member Organisation Forum on Evaluation of Publicly Funded Research*.

[51] Brutscher, P.-B., Wooding, S. and Grant, J. (2008) *Health Research Evaluation Frameworks: An International Comparison*, Santa Monica, CA: RAND Corporation, 2008.

[52] Australian Technology Network of Universities and Group of Eight (2012) *Excellence in Innovation Research Impacting Our Nation's Future – Assessing the Benefits*, Australian Technology Network of Universities, Appendix 5, p. 7.

Observations to improve the current exercise

The sectoral approach to the unit of analysis may offer certain benefits. More specifically, it may help to encourage a high level of aggregation, which is more useful in attributing impact than attempting to assess individual projects (although as noted above, this may also be the result of requesting a small number of case studies from each institution). Additionally, this approach may allow for a certain degree of flexibility in categorisation, given that one project could impact on both the society and the environment. This may help research organisations to fulfil the requirement of submitting a minimum of two case studies in three of the four sectors. However, the sectoral approach to impact may also constrain the way in which case study authors think about the impacts of their research. This raises the question of whether the research is more likely to be multi-disciplinary or multi-impact, and whether the sectoral approach to impact will negatively affect research with a diverse range of impacts.

Observations to aid scalability

The issues related to expanding the EIA trial and the unit of analysis are very much linked to the exercise's strategic intent (see section A.2.1). If the evaluation will be expanded in order to allocate funding to high-performing universities, it may be beneficial to define the research discipline as the unit of analysis – as the REF has done. However, if the purpose of the exercise is simply advocacy and accountability, the sectoral approach to impact may be more beneficial, in that it directly aligns research outputs with national priorities.

In addition, although there are only four primary impact sectors, each one encompasses many narrowly defined impact areas. For example, economic development has ten divisions, one of which is plant production and plant primary products, which in turn encompasses forestry, horticultural crops, industrial crops, summer grains and oilseeds, winter grains and oilseeds, harvesting and packing of plant products, environmentally sustainable plant production and other plant production and plant primary products. The fact that each sector covers such broad areas may pose problems in the recruitment of suitable assessment panels – particularly as there are only two assessment panels for each of the broad sectors. When the exercise is expanded it may be useful to simplify these categories and create further broad impact sectors to add to the four existing ones.

A.1.3 Describing the research-to-impact narrative

A useful heuristic in describing a research-to-impact narrative is to establish the counterfactual argument – would the impact have occurred without the research or project that is claiming the impact. This can be done in a number of ways, including through randomised control trials (RCTs) or by establishing a model of the intervention logic.[53] In the context of EIA impact case studies, however, this should be demonstrated through outlining a very clear pathway to impact throughout the case study. The guidance relating to the 'details of the impact' section (question 6) states that 'this section should provide a discussion of how the research in the following questions (7 and 8) led to the claimed impact'. However, given that this section precedes those which outline the underpinning research and its outcomes, it may be difficult to establish a

[53] European Science Foundation (2012b) *The Challenges of Impact Assessment: Working Group 2: Impact Assessment ESF Member Organisation Forum on Evaluation of Publicly Funded Research*, pp. 9–10.

logical narrative that gives clear details of how the research led to particular impacts. Indeed this problem occurred in several of the reviewed case studies (see Annex B).

The REF guidance is very similar to the EIA guidance, in that it states that impacts should be underpinned by research carried out at the institution. However, it also states that 'underpinned by' means that the research made a distinct and material contribution to the impact taking place, such that the impact would not have occurred or would have been significantly reduced without the contribution of that research.[54] In addition, the REF impact case study template follows a more chronological structure whereby a summary of the impact is outlined in section 1, followed by the underpinning research, references to the research, details of the impact and sources to corroborate the impact.[55]

Observations to improve the current exercise

Many of the reviewed case studies failed to outline adequately how the underpinning research was associated with the claimed impact. This could be addressed in several ways. As suggested above, it would be beneficial to re-order the impact case study template in order to allow for a more chronological approach to impact and research. This may aid the researcher in writing a coherent impact pathway and would reduce duplication and repetition. It may also be beneficial to add a section to the EIA guidance which states that the case study must demonstrate that the submitted underpinning research contributes to impact in such a way that the impact would not have occurred or would have been significantly reduced without the contribution of that research – as in the REF guidance. Another way in which this could be addressed is to produce a 'how to write an impact case study' handbook or provide training that details eligible impacts and how to demonstrate clearly the link between impacts and research.

Observations to aid scalability

Scalability would not be specifically exacerbated by expanding the exercise, although expansion may make the issues more problematic. Therefore the observations above should be acknowledged before rolling out the evaluation on a large scale.

A.1.4 Understanding the contribution and attribution of research to impact

There are two challenges associated with the issues of contribution and attribution. The first concerns the attribution of impact to a particular research project or institution (as outlined in sections A.1.2 and A.1.3). As noted above, this could be addressed in several ways, such as having a high level of aggregation and clearly outlining pathways to impact throughout case studies.

The second dimension concerns collaboration across multiple research institutions. Both the EIA and the REF guidelines state that 'where the submitted unit's research was part of a wider body of research that contributed to the impact (for example, where there has been research collaboration with other institutions), the case study should specify the particular contribution of the submitted unit's research and acknowledge other key research contributions'.[56] Given that there is an

[54] Research Excellence Framework (2011) *Assessment Framework and Guidance on Submissions*, REF 02.2011, p. 29.

[55] Research Excellence Framework (2011) *Assessment Framework and Guidance on Submissions*, REF 02.2011, pp. 52–3.

[56] Australian Technology Network of Universities and Group of Eight (2012) *Excellence in Innovation Research Impacting Our Nation's Future – Assessing the Benefits*, Australian Technology Network of Universities, Appendix 5, p. 17; Research Excellence Framework (2011) *Assessment Framework and Guidance on Submissions*, REF 02.2011, p. 53.

assessment for each case study within the REF and the EIA of reach – 'the spread or breadth of influence or effect on the relevant constituencies' – and significance – 'the intensity of the influence or effect' – the specific contributions of researchers should be identified. However, no criteria are provided to assess contribution. Moreover, we note that though the importance of reach and significance is outlined in the guidelines, it is not mentioned in any of the reviewed case studies (see Annex B). This suggests that more clarity and emphasis is required in determining how case studies define impact in terms of reach and significance, as well as respond to the need to address contribution.

If contribution is assessed alongside reach and significance (as outlined in the suggested scoring rubric in section 2.3) this may help to address the issue of double counting that is raised through several institutions claiming the same impact. Under the current guidance, it is implicitly the case that research impact will be split evenly across the researchers and research groups.

Observations to improve the current exercise

While reach and significance are defined as assessment criteria in the guidelines, it may be beneficial to emphasise their importance to the impact case study guidance, specifically requiring case study authors to assess their impact against these two criteria. The implementation of a new scoring rubric (as outlined in section 2.3) that includes contribution as a third assessment criterion will help alleviate the problem of double counting.

Observations to aid scalability

The problem of attribution may worsen as the exercise is undertaken on a large scale – given that the issues around double counting, attribution and contribution are more likely to occur. However, as noted above, this is only a problem if the exercise is used to allocate funding.

A. 2. Practical challenges

Practical challenges refer to those which may present problems to the operation of the exercise. In most part they are relatively minor points, which if addressed could help to improve the efficiency of the current evaluation.

A.2.1 Clarity of purpose

It is important to be clear about the strategic intent of the EIA, particularly if it hopes to link performance with budget allocations in the future.[57] At present the guidelines are somewhat vague about what the EIA Trial aims to achieve, although this is understandable given its pilot status:

> The primary purpose of the EIA Trial is to identify and demonstrate the contribution that high quality research has made to the economic, social, cultural and environmental benefit of society. Implicit in this goal is the purpose to investigate the means by which these benefits may best be recognised, portrayed and assessed by institutions and government.[58]

[57] Grant, J. et al. (2010) *Capturing Research Impacts: A Review of International Practice*, Santa Monica, CA: RAND Corporation, pp. 62–3.

[58] Australian Technology Network of Universities and Group of Eight (2012) *Excellence in Innovation Research Impacting Our Nation's Future – Assessing the Benefits*, Australian Technology Network of Universities, p. 3.

However, the guidance does note that on completion a consolidated assessment report will be prepared for each institution, a summary report of results spanning all institutions will be made available to each participating institution and a summary EIA Trial report will be prepared for the peak EIA Governance Group and released to the public following endorsement – which will evaluate the process of the EIA trial and make recommendations concerning its potential for more widespread adoption. If widespread adoption will be linked to budget allocations, several points should be considered, which are outlined below in relation to scalability.

The REF guidance is very clear concerning the above-mentioned points. It states that one of its key purposes is for the four higher education funding bodies conducting REF to use the assessment outcomes to inform the allocation of their grant for research to the institutions which they fund, with effect from 2015–16. It goes on to state that the criteria and procedures that will be applied in the assessment will be published in full, well in advance of institutions making their submissions. Moreover the outcomes will be published in full and decisionmaking processes will be explained openly. The primary outcome of the REF will be an overall quality profile for each submission, judged on quality of research, which will inform the ultimate funding decisions.

Observations to improve the current exercise

Although the output of the EIA trial is not entirely apparent in the guidance, it is clear that this is an exploratory exercise, so whether the evaluation will lead to funding decisions has yet to be determined. If this does become the case, the observations to aid scalability should be considered.

Observations to aid scalability

If the EIA will be linked to funding decisions in the future, the guidance should be explicit in how performance is related to funding at the outset and the approach should provide a system for the unambiguous rating and comparison of submissions as well as addressing this issue of potential double counting as discussed above. Moreover, the relationship between the assessment of impacts and the allocation of funding should be considered (this could be a relationship where all assessed impacts are awarded or one that aims to concentrate funding in areas of excellence).[59] Given that the REF has already addressed these points, it could provide a useful starting point for developing the EIA guidelines.

A.2.2 Definitions

It is essential to have a clear definition of impact when assessing impacts across various areas and institutions. This helps to avoid problems such as researchers putting forward impacts which the EIA does not think are valuable, difficulties in comparing different researchers or research groups using various definitions, and researchers defining and valuing achieved impacts after the research has been completed.[60]

The definitions employed in the EIA and REF guidance are largely the same. Impact is defined in both documents as an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life beyond academia:

[59] Grant, J. et al. (2010) *Capturing Research Impacts: A Review of International Practice*, Santa Monica, CA: RAND Corporation.

[60] Grant, J. et al. (2010) *Capturing Research Impacts: A Review of International Practice*, Santa Monica, CA: RAND Corporation, p. 63.

> Impact *includes*, but is not limited to, an effect on, change or benefit to:
>
> - the activity, attitude, awareness, behaviour, capacity, opportunity, performance, policy, practice, process or understanding
> - of an audience, beneficiary, community, constituency, organisation or individuals
> - in any geographic location whether locally, regionally, nationally or internationally.
>
> Impact *includes* the reduction or prevention of negative effects including the harm, risk or cost arising from negative effects.[61]

The EIA guidance goes on to state:

> It does not include impact on research or the advancement of academic knowledge, nor impacts on students, teaching or other activities within the submitting institution. It may include impacts within the higher education sector, including teaching or students where they extend significantly beyond the submitting higher education institution.[62]

This is slightly adapted from the REF guidance, although the meaning does not differ across the two documents.

The definition and conceptualisation of impact has faced criticism from the academic community. Smith et al. outline how many feel it threatens researchers' autonomy and academic freedoms, and there is also concern regarding the broad scope of 'impact' and the various ways in which it could be interpreted by the academic community. There is also concern that impact of the REF has been specifically aligned with national priorities, and this may 'disadvantage researchers whom they fund should the specific social, economic and cultural benefits that charities, for example, seek to promote fail to coincide with national priorities'.[63]

The definition of research differs in formulation across the EIA trial and the REF guidance, although they both agree on the fundamental points. A fundamental difference between the REF and the EIA guidance is clarity around how quality research should be defined. Both the REF and the EIA state that impacts should be based on quality research, although unlike the REF (which states that research must be of 'quality that is recognised internationally in terms of originality, significance and rigour'[64]) the EIA does not set a quality threshold. This may lead to confusion concerning how the quality of research should be measured or outlined in the case studies.

Observations to improve the current exercise

Setting out a clear definition of impact and research in the guidance is very useful, and the EIA does this well. One small operational recommendation to consider is to include the definitions of impact and research in section 2.4.1 of the guidance, which provides question-by-question

---

[61] Research Excellence Framework (2011) *Assessment Framework and Guidance on Submissions*, REF 02.2011, p. 26 (emphasis in the original).

[62] Australian Technology Network of Universities and Group of Eight (2012) *Excellence in Innovation Research Impacting Our Nation's Future – Assessing the Benefits*, Australian Technology Network of Universities, Appendix 5, p. 12.

[63] Smith, S., Ward, V. and House, A. (2011) '"Impact" in the proposals for the UK's Research Excellence Framework: shifting the boundaries of academic autonomy', *Research Policy*, 40(10), pp. 1369–79.

[64] Research Excellence Framework (2011) *Assessment Framework and Guidance on Submissions*, REF 02.2011, p. 43.

guidelines on how to complete the impact case study template. This would provide a quick point of reference for researchers when consulting the guidance.

Observations to aid scalability

The definitions have been well laid out and therefore scalability would not pose a problem. No further action relating to definitions is required before expanding the exercise.

A.2.3 End users

An important issue in qualitative assessments of impact is the need to engage end users of research. In the EIA exercise, it is essential that each review panel is made up of reviewers who can make an independent assessment of high impact research versus low impact research. This means recruiting end users of research such as business people, policymakers and others.[65] The EIA guidance states that the two assessment panels for each broad sector will comprise 'a mix of end-user representatives with expertise in the broad range of disciplines represented by the cluster and academic staff with disciplinary expertise. The end users will constitute the majority of the Assessment Panel and shall also provide the Chair for each Assessment Panel.'[66] This differs considerably from the composition of the REF assessment panels. Each REF sub-panel is predominantly made up of 'practicing researchers of suitable personal standing who collectively have an appropriate breadth of research expertise and carry the confidence of the community', although sub-panels 'should normally also include members from the private, public or third sectors with expertise in commissioning, applying or making use of research'.[67] With this in mind, the EIA trial needs to engage with end users considerably more than the REF in order to recruit assessment panels.

However, engagement with end users is not only relevant at the end of the process. In order to produce impact case studies both the EIA trial and the REF assume some level of engagement with end users. The EIA guidance states that each impact case study should contain the names of people who have made a primary contribution to the case study, which could include end users or beneficiaries of the research.[68] The REF goes a step further. The guidance for each main panel (which differ slightly) states that institutions must outline their 'approach to impact'. The Main Panel A guidance states institutions should 'describe the unit's approach to interacting with non-academic users, beneficiaries or audiences and to achieving impacts from its research, during the period 2008–2013'.[69] Therefore institutions should be engaging with end users throughout the process and this may have secondary implications if the exercise is undertaken on a larger scale.

---

[65] Grant, J. et al. (2010) *Capturing Research Impacts: A Review of International Practice*, Santa Monica, CA: RAND Corporation, p. 68.

[66] Australian Technology Network of Universities and Group of Eight (2012) *Excellence in Innovation Research Impacting Our Nation's Future – Assessing the Benefits*, Australian Technology Network of Universities, Appendix 5, p. 11

[67] Research Excellence Framework (2010) *Units of Assessment and Recruitment of Expert Panels*, REF 01.2010, p. 5.

[68] Australian Technology Network of Universities and Group of Eight (2012) *Excellence in Innovation Research Impacting Our Nation's Future – Assessing the Benefits*, Australian Technology Network of Universities, Appendix 5, p. 9.

[69] Research Excellence Framework (2012) *Panel Criteria and Working Methods*, REF 01.2012, p. 33.

Observations to improve the current exercise

Composing an assessment panel that is primarily made up of end users has not been done before and is commendable. However, the engagement of end users in panels and throughout the process may become more problematic when the exercise is undertaken on a large scale. Therefore the observations to aid scalability noted below should be considered.

Observations to aid scalability

The EIA intends to use two assessment panels for each of the four broad sectors. As the four sectors encompass a very wide range of subject areas (as outlined in section A.1.2) the assessment panels necessarily need to be very diverse, and each assessment panel will have a significant workload (it is worthy of note that the REF employs four main panels and 36 expert sub-panels – although the REF is a larger undertaking than the EIA). In addition to end-users' participation in the assessment panels, end users may also be required to interact with institutions throughout the case study writing process, in providing evidence of impact. With this in mind, in order to recruit and maintain the necessary people for each panel, as well as ensure sustained end-user engagement, it may be necessary to identify the motivation of the experts in order to incentivise them if engagement is low in the future.

Annex B : Case study review

B. 1. Methodology

The purpose of the case study review was to focus on the overall research-to-impact narrative and how this followed from the guidelines. In particular, we aimed to identify strengths and weaknesses across the case studies, which would allow us to determine how impacts were being identified and presented in the case studies, and whether any common strengths and weaknesses emerged which would give us insight into how to improve future impact assessment exercises and provide guidance to case study authors and institutions.

From the 64 case studies submitted between the five universities in the ATN for the EIA trial, ten case studies were selected for review within this study. The selection ensured there was a mix of universities, a range of panels and a range of scores. This resulted in the review of:

- two case studies per submitting university
- case studies covering all four panels:
    - one from the Defence Panel
    - three from the Economic Development Panel
    - four from the Society Panel
    - three from the Environmental Development Panel
- case studies covering the different levels of scores given by the panels:
    - two 'A's
    - two 'B's
    - two 'C's
    - two 'D's or 'E's
    - two unrated.[70]

Where there was more than one case study of a certain score within a university and panel a case study was selected randomly.

Each case study was reviewed by two people and discussed at a cross-cutting workshop attended by the whole team, when underlying themes and trends across all of the case studies were developed and identified. These are broadly divided into three areas:

- observations on the overall process within the EIA trial

[70] This occurred where society sub-panel 1 ran out of time to complete the assessment.

- ideas to improve the case studies within the current guidelines
- amendments to the process which could improve the case studies produced.

Each is discussed in more detail below. In order to illustrate the themes, we have tried where possible to provide examples from the case studies we reviewed, though were sometimes unable to give details because of the need to preserve confidentiality of those involved.

B. 2. Observations on the overall process

Three themes came across as overall points that the EIA advisory board may want to consider before repeating this exercise: prior knowledge, action or implementation research, and size of case study.

Prior knowledge

In one case study that we reviewed, the comments from the panel suggested that they had detailed knowledge of the work, which assisted them in awarding a high score. However, it did not appear to us that the case study as written necessarily merited this high score as the impact narrative was not clear and not suitably evidenced. It is possible that other examples of this existed in case studies we did not review. To ensure that the process is fair and transparent it is important that case studies are based solely on the information within them rather than additional information and insight the panel may have into a project. To eliminate this bias, thought needs to be given to the conditions under which panel members assess case studies about which they have detailed knowledge. We recognise this will be a difficult balancing act, as clearly there is merit in having expert knowledge of the underpinning research and impact, but this must be tempered to ensure the scoring process is fair and balanced.

Action or implementation research

Action or implementation research is research conducted in the society setting in which it aims to have a direct effect. Action research, in particular, is a specific methodological approach whose hypotheses and methods may be shaped in a reflexive manner as the research is undertaken. In other words, the actions of the researchers themselves may influence the nature of the findings because they are actively engaged in both research and delivering a given intervention. Using this as an underlying theory, academics and practitioners can work together to implement solutions that work in a given environment.

While this is a very useful research approach, it can be problematic for impact case studies because it is difficult to see where the research ends and the impact begins, as the impact is designed as an outcome of the research project. For example, in one case study we examined, a policy review led to changes being implemented, which themselves were then further evaluated by the same researchers. Two of the case studies out of the ten we reviewed seemed to contain some element of action or implementation-based research, so this is likely not to be an isolated problem. Therefore, the EIA board may want to consider how they assess or permit action- or implementation-based research to be presented and make this explicit in the guidance. In addition, care needs to be taken to ensure that researchers with less premeditated examples of impact are not deterred from submitting.

Size of case study

The case studies we reviewed covered a broad body of work, often describing the research across a whole institution as opposed to single research areas or suites of research projects. This may be an

artefact of the selection process, which asks universities to produce a minimum of two and a maximum of five case studies per sector, as opposed to the UK REF system, which links the submission to the number of staff, requiring one case study per ten staff submitted, and the use of SEO codes rather than academic disciplines, which encourages cross-institutional multi-disciplinary research.

While noting that we only reviewed ten case studies, there are nevertheless a number of implications that this broader approach to the relative 'size' of the case study may have. First, of the case studies we reviewed, the panel appears to have awarded higher scores for broader bodies of work undertaken by entire institutes, rather than pieces submitted by individual researchers or small research groups. In three cases, this institutional approach strengthened the case study as the larger body of work undertaken by the institute led to a broader set of impacts which could be claimed. In addition, the impact narratives in these cases were clear and coherent, but in one case we felt it over complicated the narrative as it brought together the individual research interests of a group of researchers and these different strands of research were difficult to disentangle. As a result, it was not a cohesive case study and did not present a clear set of impacts. These relative strengths and weaknesses may need to be looked at more closely in future.

One possible reason for this tendency to include broader research areas in one case study is the sectoral approach of the EIA. The EIA trial was conducted across four sectoral categories: defence, society, economic development and environmental. These categories are very broad and though the authors are able to indicate secondary sectors to which the impact of the research is relevant, the sectoral and broader approach to *impact,* as opposed to a disciplinary approach to *research,* may have the effect of encouraging researchers to look across many disciplines for impacts occurring in one area.

However, the converse of this is that it may also discourage researchers from considering the multiple types of impacts, across sectors, their research has had. In other words, an impact trial such as the EIA can ask for impact to be presented in one of two ways. First is the approach taken by the REF, which takes the discipline of the research and considers the multiple types of impact which resulted. Second is the approach taken by the EIA, which starts with a single type of impact and asks for consideration of the potentially multiple types of research that contributed to it. Both approaches call for some sort of interdisciplinary effort. The REF actively encourages more interdisciplinary impacts to be considered, whereas it is our view that the approach of the EIA discourages this, yet it does encourage more multi or interdisciplinary research to be included in a single case study. This represents a strategic trade-off for the EIA to think through in the future.

B. 3. Ideas to improve the case studies within the current guidelines

The current guidelines set out how authors should prepare their case studies and provides guidance to the panel on criteria for assessing the case studies. However, several issues arose across numerous case studies, which suggest the guidance may need revisions for future iterations: justification of the impact, and reference to potential versus future impacts, word limits, use of the context section, and demonstration of research quality.

1. Justifying and providing evidence for the impact

In five out of ten cases, our review team felt that the case study lacked evidence to confirm the stated impacts. The nature of this lack of evidence varied. For example, in one case study we felt that impacts were overstated and not fully backed up with a corresponding breadth of supporting evidence that could be verified. In two other cases, academic and societal impacts were conflated. For example, the authors claimed impact from the presentation of research findings and/or the publishing of journal articles or books. This highlights the need to give clear guidelines about what will count as impact (see Section A.2.2). It also draws attention to the need to be clear about how to evidence impact and the potential need to train researchers in preparation for the assessment. This will help to ensure they can recognise and articulate impact from the initial design stage of a research project to its completion and dissemination.

2. Potential impacts

In three cases, the authors claimed future, potential impacts and it appears to us that this was not always picked up by the assessment panels. For example, the authors of one case study listed the expected value of economic gain from their research and this was accepted by the scoring panel and cited as evidence of impact. The guidelines state that impact must have occurred in the stated census period (between 1 January 2007 and the date of submission) and it is important that authors and assessors only consider impacts, and evidence for those impacts, within those dates.

3. Word limits

Many of the case studies exceeded the indicative word limits for some of the sections, but all the case studies reviewed were within the 15 page limit. The EIA advisory board may want to consider reducing the overall length of the case study submissions if scaling the assessment up to review the impact of research across all Australian universities, as this increase in volume will create a significant burden on the assessors on the panels.

4. Use of the context section

The guidance states that the context section should be used to provide relevant background information about the nature of the research problem. We felt that this section could significantly improve the overall case study. An up-front understanding of why the topic and research is relevant and any needed background information is potentially useful for the reader in helping to frame the research problem and the context for the argument of impact. However, we observed that in all but three cases (out of ten case studies reviewed), the section was not correctly used. For example, some case studies used it to provide additional detail about the background of the researchers and institution involved, or used it as a summary of the case study, which weakens the summary section or makes it repetitive. This contravenes the guidance, which states that the information in the summary should not repeat any information given elsewhere in the template. The EIA board may consider re-wording the guidance to ensure that authors outline the nature of the problem and why research is required, and do not use it to repeat things said in other areas of the case study. In addition, exemplar case studies with well-written context sections could be provided to assist researchers in drafting this section and understanding what should be included.

5. Demonstrating research quality

The guidelines state that 'evidence about the quality of the research' should be provided in order to assess the relationship between research quality and impact, but no definition of quality is

stipulated nor ways for demonstrating it provided. Where journal articles are listed as outputs this can be assessed by other members of the research community on the panel, but where lay people are involved or outputs other than scholarly articles are provided, additional guidance may be required to enable the panel to make an informed decision about the quality of the research, and to assist the institutions in determining what to submit. Four of the ten case studies demonstrated the quality of their research through descriptions, download figures, impact factors and Google and Scopus cites.

B. 4. Amendments to the process which could improve the case studies produced

There were several areas in which additional guidance from the EIA advisory board would improve the process through which case studies were produced, and subsequently the quality of case studies received: attribution, impact pathways, cross referencing, timelines and the breadth and significance of impact.

1. Contribution

None of the case studies we reviewed stated there had been collaborations or contributions from others outside the institution claiming the impact, but increasingly, research is conducted as a collaborative effort. In a trial such as this it is not an issue, but if the assessment was to be expanded across Australia, it would have to be addressed as multiple institutions could be claiming the same research and impact and the EIA advisory board would need to provide guidance on how scores would be awarded to contributing parties. One way this could be achieved would be to require the case study to state the specific contribution of the submitting institution in the collaborative project. This might also be included as an explicit assessment criterion in addition to breadth and significance.

2. Impact pathways

The production of impact from academic research can be visualised as a pathway that involves various steps: inputs, processes, outputs and outcomes, and impact, as illustrated in Figure B.1. Consideration of research in this way can assist researchers in understanding and articulating the relationships between input, process and output in their research and help them narrate their case study in a logical order.

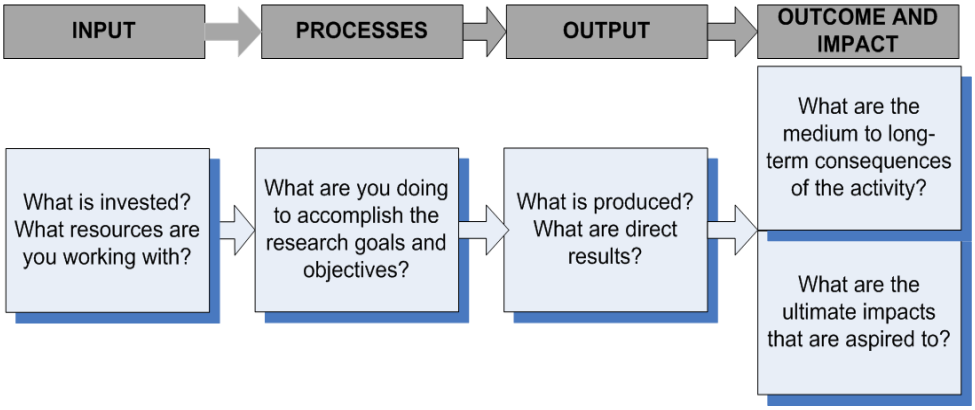

Figure B.1: The impact pathway

In the majority of case studies we reviewed, the pathway from academic research to wider impact was not clearly articulated. For example, in some cases primary research was stated as having

implications for policymaking, without detailing the process by which the policymakers were alerted to and engaged with the research outputs and researcher. One reason for this could be the non-chronological ordering of the template sections, in which the impact section is required before the description of the underpinning research. In many cases the author introduced elements of the research in the impact section in order to provide the context that the reader does not already have.

In addition, often the impact pathway needs to be 'pushed' further. In our review we observed that case study authors seemed only to comment on the impact of their research on the people with whom they directly engage, but there were many examples where it seemed to us that impacts could be found beyond this immediate group. Authors need to look further afield and understand and explain how the organisations and individuals they have interacted with use their research outputs and take them further, involving new user groups and generating new impacts.

One way of improving the articulation of the impact pathways is to consider the counterfactual: what would *not* have happened if the research had not been conducted. More guidance from the EIA and training of the institutions and case study authors is required to assist the authors in understanding impact, developing the chain of evidence, and considering the full societal reach of their impact through to the final end user.

3. Cross referencing

The guidance does not mention the use of cross referencing between sections, but it is a helpful technique. One case study used this to interlink the research outputs to the impacts claimed. In addition, only two out of the ten case studies provided explanations of the web-links provided in the validation of impact section. These are useful to remind the reader of the relevant impact or research which the source supports. The EIA advisory board may consider adding these to the guidance, especially as the evidence section does not follow immediately from the impact section in the current impact template.

4. Timelines

Six case studies did not mention the timelines within which the impact took place. This makes it unclear when things occurred and therefore whether they are eligible within the assessment criteria, and could lead to a case study being incorrectly categorised as unclassified. The EIA advisory board may like to add to the guidelines to ensure that authors state dates related to the projects and impacts they described.

5. Breadth and significance of impact

The author guidance states that the discussion around impact 'should further outline the reach and significance of the impact' and the panel guidelines state that the 'panel will assess each case study against overall reach and significance rather than assessing each component separately', and goes on to specify the scale for assigning a rating. However, none of the case studies that we reviewed referred to these terms, and instead required the panel to draw out these elements from the information provided. This raises questions for us about how these judgements were made by the assessment panel and on what basis. Thus, we suggest that the EIA advisory board might consider adding to the guidance to state that these terms should be used in the impact section so that there is a clear and common basis against which assessment panels make judgements about the extent of breadth and significance claimed, and the evidence for it.

Annex C : Analysis of panel scoring data from the ATN universities that participated in the EIA trial

C. 1. Summary

The primary purpose of the analysis of panel scoring data from the EIA trial for the ATN universities was to determine how the individual scores of panel members influenced their draft impact ratings and whether the panel group meetings resulted in major 'shifts' or 'deviations' within the panel that led to changed final impact category ratings assigned to case studies. In order to determine this we conducted a quantitative assessment of the panel scoring data contained in the reviewer scoring sheets and analysed the data for all the ATN university case studies.

C.1.1 Main messages

As much of the data and underpinning analysis is complex, we will begin by providing a summary of the main messages from the analysis:

- In general, the intra-panel reliability was good: there were relatively good levels of consensus between the individual 'assessor draft ratings' and the final impact category ratings assigned to the different case studies (overall as well as per panel) (see Section C. 4).
- The final impact category ratings of the case studies that were reviewed by both sub-panels within their respective panels indicated good levels of inter-panel reliability (see Section C.3.1).
- Across all the panels, the 'scores' provided in sections 2, 3 and 4 of the scoring sheets were significantly correlated with the corresponding 'assessor draft ratings' assigned by the reviewers, with the score in section 3 (the 'rating of description of the nature and extent of the impact') having the most influence on the draft rating (see Section C. 5).
- When assessing the levels of burden involved in assessing impact case studies, it was found that the average assessment time per case study for each reviewer was 28.4 minutes (across all four panels) and, in general, there were no major differences in the average assessment times across panels (see Section C. 4). Additionally, across all the panels, the average total time spent on assessing each case study was 112.1 minutes.

- The overall distribution of the ATN university case study impact category ratings illustrates that almost 80% of the ratings were classified as either 'A' ('outstanding impact in terms of reach and significance'), 'B' ('very considerable impact in terms of reach and significance') or 'C' ('considerable impacts in terms of their reach and significance') (see Section C.3.1). The distribution of case study ratings shows that panel members were able to distinguish and differentiate between impact case studies.

C. 2. Introduction

The quantitative analysis of the scoring data presented in this annex focuses on data from the five ATN universities that participated in the EIA trial. The analysis was carried out from the following two perspectives:

- Analysis of the final case study ratings to obtain an overview of how the ATN universities performed as a whole (a 'macro' perspective); this was broken down into a panel-level assessment and a university-level assessment.
- Analysis of the data from the individual scoring sheets (assessor draft ratings, scoring data, time taken to complete assessment) submitted by each reviewer (a 'micro' perspective); the primary aims of this analysis were to obtain, for each panel, a general idea about the levels of consensus between the individual draft ratings and the final impact ratings awarded to the case studies (the intra-panel reliability), and to determine whether some of the individual scores assigned by reviewers influenced their overall draft ratings (and if so, to what degree). In addition, to obtain some idea of the burden involved in carrying out the assessments, the reviewer assessment times were analysed.

C.2.1 Methodology

The following fields of data in the scoring sheet template provided to each reviewer have been quantitatively analysed in this annex:

- 'overall assessor draft rating of case study (A–E)' (also referred to in the text as individual draft rating or simply draft rating)
- scoring (1–5)
    - 'rating of description of how the research underpinned the impact'
    - 'rating of description of the nature and extent of the impact'
    - 'rating of the validation of the impact'
- time to complete the assessment: 'How long did it take you to complete the assessment for this case study?'

Additionally, the final 'letter' impact category ratings that were awarded to the various case studies in each panel have been included in some of the analysis presented in this annex.

In order to carry out parts of the analysis, the 'letter' ratings assigned by reviewers and/or panels were given numerical equivalents, as listed in Table C.1. Whenever this conversion scheme has been used in the analysis, it has been explicitly stated in the associated text.

Table C.1: Numerical equivalents of 'letter' ratings used in some of the analysis

Letter rating	Numerical rating
Not classified (NC)	0
E	1
D	2
C	3
B	4
A	5

In the examination of the scoring data, it was noted that the number of reviewers whose individual scoring sheets were provided to us did not tally with the list of reviewers included in the EIA trial report (some reviewers did not send their scoring sheets in advance but brought them to the panel meetings to inform their input in panel discussions). The analysis presented here uses the scoring sheet data that were made available to us. When the data were collated from the scoring sheets, it was observed that some reviewers deviated from the standard A, B, C, D or E draft rating structure prescribed in the template and included ratings such as B+, B−, C+, C−, 5, 4, 3 or 2. In our analysis, we have approximated these individual draft ratings (21 occurrences) as follows: draft ratings of 5, 4, 3 or 2 were assigned ratings of A, B, C or D, while ratings with the structure B+, B−, C+ or C− were approximated to B or C. In addition, six case studies were given more than one draft rating and in these instances we have taken the higher rating for the purposes of our analysis.

Caveats

There are some important qualifications that must be taken into consideration while interpreting the results presented here. Since the analysis has focused on data from the five ATN universities, it would not be appropriate to extend the interpretation to the whole EIA trial. By including data from only a subset of the universities that participated in the trial (the ATN universities), we might not be able, for example, to obtain an accurate view of the overall scoring 'nature' of the reviewers within a particular panel. Furthermore, some of the findings (for example, those related to the examination of data from the Defence Panel) will have the caveat that they are based on relatively small amounts of data. Nevertheless, the analysis presented below may be used to provide a general indication of various characteristics of the scoring behaviour in the context of the ATN universities.

C. 3. Analysis of case study impact category final ratings

C.3.1 University-level analysis

A total of 64 case studies were submitted for assessment. The distribution of case studies across the five ATN universities has been listed in Table C.2. With 14 case studies each, Institution B and Institution C had the maximum number of submissions, and Institution D the least (11). Seven of the 64 cases studies were sent to both sub-panels within their respective panels for

assessment (three economic development case studies, two environment case studies, and two society case studies). Furthermore, of the 64 case studies, seven were not reviewed by the Society Sub-Panel 1, whose members ran out of time to complete their assessment (these cases have been specified as 'unrated' case studies in this analysis). A further case study was sent to the wrong panel (Society 2) and was also therefore classified as unrated. However, the unrated case studies did receive individual assessor ratings.

Table C.2: Number and percentage distribution of case study ratings for each ATN university

Institution	Number of case studies	Distribution of case studies across the universities (%)
Institution A	12	19
Institution B	14	22
Institution C	14	22
Institution D	11	17
Institution E	13	20

To obtain a general idea of how each university performed in the trial, Figure C.1 shows the distribution of case study ratings across the ATN institution. Institution B was awarded the most (3) 'A' ratings ('Outstanding impact in terms of reach and significance'). Institutions C, D and E did not receive any 'A' ratings while Institution A was given one 'A' rating. Along with Institution A, Institution B also received the most (5) 'B' ratings ('Very considerable impacts in terms of their reach and significance'). Institution C was awarded the most (11) 'C' ratings ('Considerable impacts in terms of their reach and significance'), followed by Institution A (6). Institutions B, D and E all received three 'D' ratings ('Recognised but modest impacts in terms of their reach and significance'). Overall, there was only one 'E' rating ('The research impact is of limited reach or significance'), which was awarded to a case study submitted to Environment Sub-Panel 2. The only 'not classified' (NC) rating was given to an Economic Development Sub-Panel 2 case study.

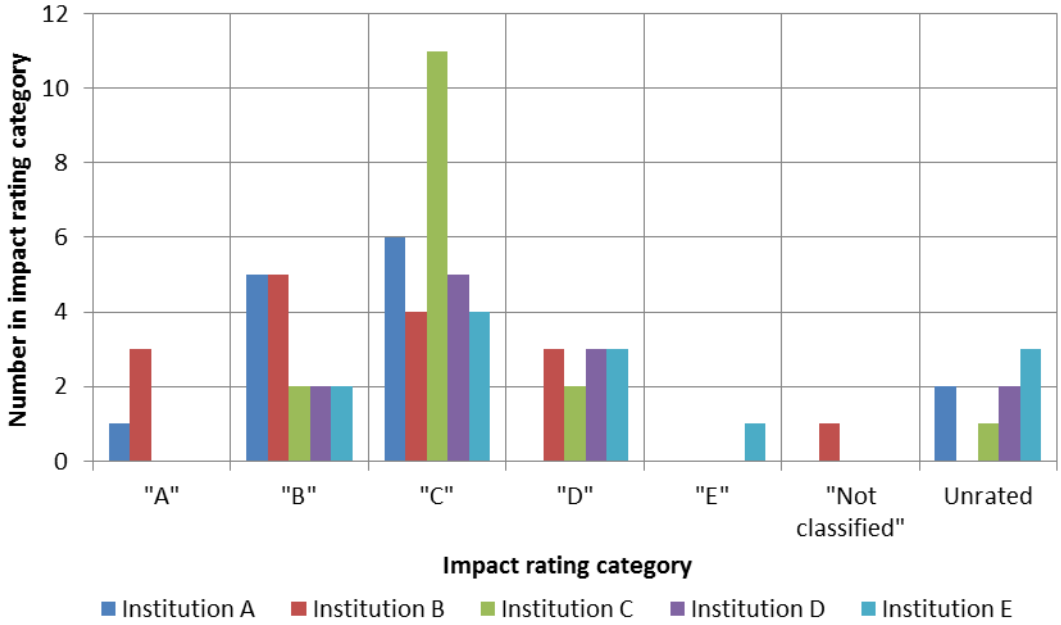

Figure C.1: Distribution of case study impact category ratings (in numbers) for each ATN university

The distribution of case study ratings (expressed as a percentage and disregarding those that were unrated) for the five universities is listed in Table C.3. For each of the universities, the impact category rating entry in the table with the maximum occurrence frequency (%) has been highlighted in bold. Except for Institution B, which received a maximum share of 'B' ratings (approximately 31% of all their ratings), all the other institutions received the most number of 'C' ratings, with Institution C receiving as many as 11 (around 73%) 'C' impact category ratings.

Table C.3: Percentage distribution of case study impact category ratings for each ATN university

Institution	Percentage of 'A' ratings	Percentage of 'B' ratings	Percentage of 'C' ratings	Percentage of 'D' ratings	Percentage of 'E' ratings	Percentage of 'NC' ratings
Institution A	8	42	50	0	0	0
Institution B	19	31	25	19	0	6
Institution C	0	13	73	13	0	0
Institution D	0	20	50	30	0	0
Institution E	0	20	40	30	10	0
All universities	6	25	48	17	2	2

For comparison, as well as to obtain a consolidated view of the performance of the ATN institutions, the distribution of case study ratings (excluding those that were unrated) for all the universities are also presented in Table C.3 and the distribution is shown in Figure C.2. The fact that there is a distribution of ratings shows that panel members were able to distinguish and differentiate between impact case studies during their assessment. As illustrated, across the trial, almost 80% of the ratings were classified as 'C' or higher, and around a third of the ratings were

judged by their respective expert panels to be either 'outstanding' ('A') or 'very considerable' ('B') in terms of the reach and significance of the impact presented in the corresponding case studies.

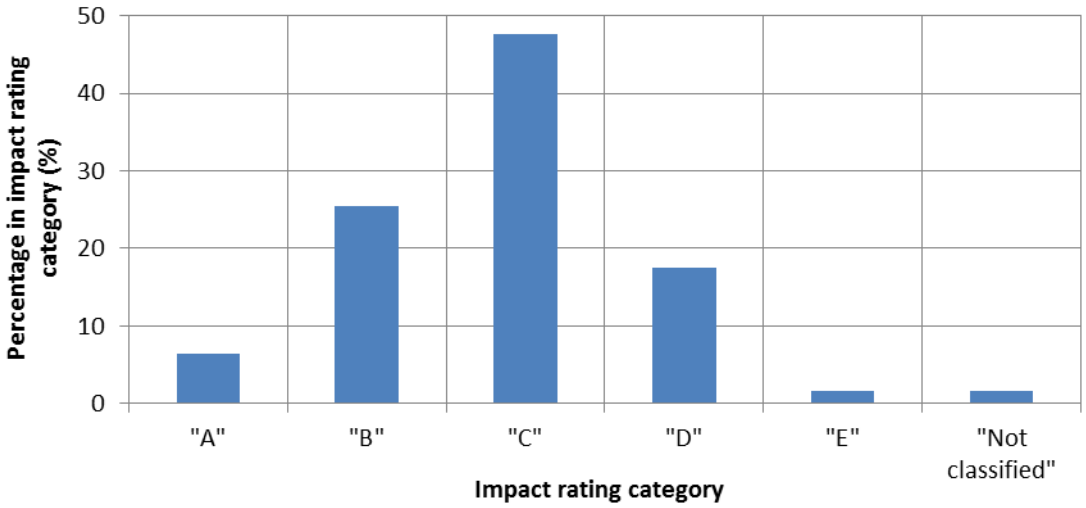

Figure C.2: Overall distribution of case study impact category ratings

The numerical equivalents of the 'letter' ratings (listed in Table C.1), excluding unrated case studies, were added up for each university to yield an aggregate 'number' score, and an average 'number' score was then calculated. These have been listed in Table C.4, which shows that Institution A (with an average 'number' score of 3.58) was the 'best' performing university (Institution A received no 'D', 'E' or 'NC' ratings). Institution B followed closely behind Institution A obtaining an average 'number' score of 3.31 for its 16 ratings. The ten ratings awarded to Institution E resulted in the lowest average 'number' score (2.70) of the five ATN universities.

Table C.4: Aggregate 'number' scores and corresponding average 'number' scores for each ATN university

Institution	Number of ratings	Aggregate 'number' score	Average 'number' score
Institution A	12	43	3.58
Institution B	16	53	3.31
Institution C	15	45	3.00
Institution D	10	29	2.90
Institution E	10	27	2.70

C.3.2 Panel-level analysis

The final case study ratings were also examined from the point of view of their respective panels. The occurrence frequency histograms of the various ratings (excluding those that were unrated) for each of the four main panels have been plotted in Figure C.3. Of the four Defence Panel ratings, two were rated as 'D' (50%) and one each were assigned ratings of 'B' (25%) and 'C' (25%). In the Economic Development Panel, 20 of the 27 impact ratings (approximately 74%)

were an 'A', 'B' or a 'C'. All of the Society Panel ratings (17) were rated at least 'C'. In the Environment Panel, 11 of the 15 ratings (approximately 73%), were rated at least 'C'.

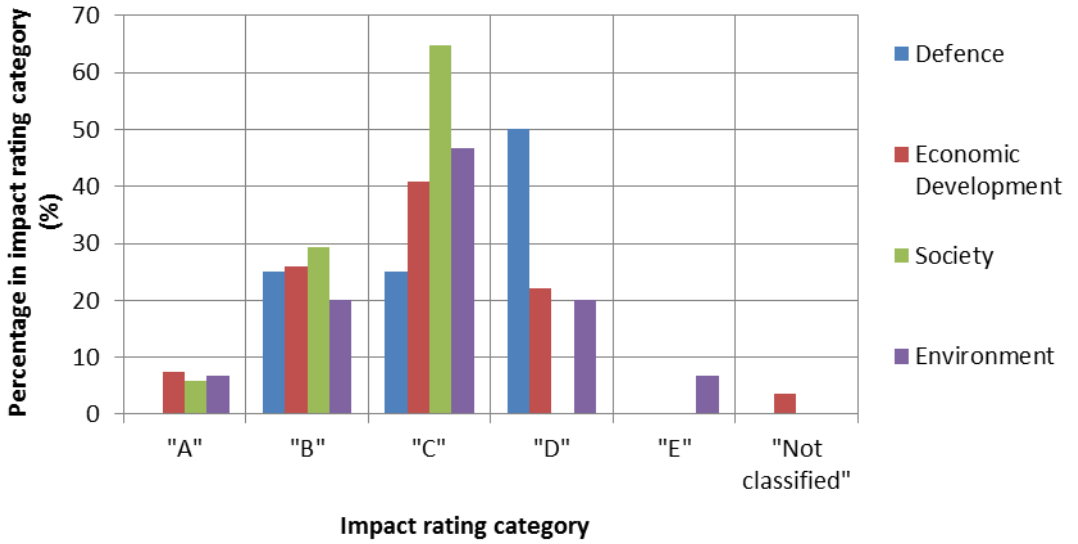

Figure C.3: Overall distribution of case study impact category ratings across the four main expert panels

As with the individual universities (Table C.4), each rating in a given panel was assigned an equivalent numerical score (according to the scheme listed in Table C.1 and disregarding those that were unrated). On the basis of the number of ratings allocated by each panel (Defence Panel: 4; Economic Development Panel: 27; Society Panel: 17; Environment Panel: 15), an average 'number' score was calculated for each panel. These are presented in Table C.5 and demonstrate that, on average, the Society Panel received the highest score (3.41) for their 17 ratings. The Society Panel was followed by the Economic Development, Environment and Defence panels, which received average 'number' scores of 3.07, 3.00 and 2.75, respectively. When the average 'number' score was approximated to the nearest whole number and then assigned an equivalent 'letter' rating (according to Table C.1), all four panels were classified as 'C' (having case studies whose reach and significance exhibited 'considerable' impact).

Table C.5: Average 'number' scores, rounded average 'number' scores and overall 'letter' ratings for each of the four panels

Panel	Number of ratings	Average 'number' score	Rounded average 'number' score	Overall 'letter' rating
Defence	4	2.75	3	C
Economic Development	27	3.07	3	C
Society	17	3.41	3	C
Environment	15	3.00	3	C

Inter-panel reliability

As mentioned previously, seven of the 64 case studies were submitted to both sub-panels within their respective panels resulting in 71 ratings awarded overall (including those that were unrated). All but one of these case studies received dual ratings and an examination of these ratings

demonstrated that there were good levels of inter-panel reliability. Specifically, one of the case studies that was assessed by both Society sub-panels received a 'C' rating from sub-panel 2 but was unrated by sub-panel 1 (whose members ran out of time to complete the evaluation). Of the remaining six case studies, two received identical ratings from both sub-panels, three received ratings that were within one rating point of each other, and one obtained ratings that differed by two rating points. (One case study in the Environment Panel received a rating 'between C and D' from sub-panel 1 and a rating of 'D' from sub-panel 2; for the analysis presented here, the rating assigned by sub-panel 1 was taken to be a 'C'.)

C. 4. Analysis of assessor draft ratings and assessment times from individual scoring sheets

In this section, data from the individual scoring sheets submitted by each reviewer have been examined. In particular, the assessor draft ratings (section 1 of the scoring sheet) for each case study have been compared to the final impact category ratings to determine whether or not there were any major deviations or shifts within a given panel after the panel meetings, to highlight levels of consensus within a given panel. In addition, in order to gain an insight into the levels of burden involved in assessing the cases studies, the time taken by reviewers to assess case studies has also been analysed (section 8 of the scoring sheet). The data for the various expert panels and the results from each of the panels have been collated and presented in turn. First the generic process used in the analysis to demonstrate intra-panel reliability has been described below (using data from one of the expert panels as an example).

C.4.1 Methodology employed to demonstrate intra-panel reliability

The first step in examing the degree of consensus between the individual draft ratings and the corresponding final rating awarded to a particular case study involved the determination of the distribution of draft ratings for each case study. As an example, the individual draft rating distribution for one of the case studies that received six draft ratings is depicted in Figure C.4. As a visual indication, the 'position' of the final rating that was awarded by the panel to this particular case study is also highlighted in the figure.

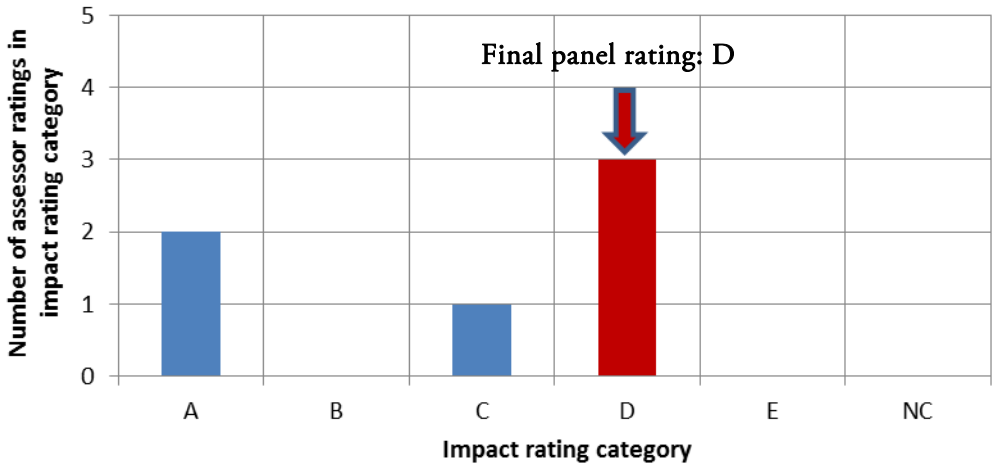

Figure C.4: Distribution of individual assessor ratings for an example case study

Next, for each case study, the distribution of the deviations of the individual draft ratings from the final rating was calculated. The following four categories of deviation were determined:

- 0 deviation: the number of instances where the individual draft ratings were identical to the final case study rating
- deviation of ±1: the number of instances where the individual draft ratings differed from the final rating by one 'rating point'
- deviation of ±2: the number of instances where the individual draft ratings differed from the final rating by two 'rating points'
- deviation of ±3 or more: the number of instances where the individual draft ratings differed from the final rating by three or more 'rating points'.

The greater the number of 0 deviations, the greater is the degree of consensus between the individual draft ratings and the final rating. Conversely, a greater occurrence frequency of ±3 deviations would indicate that there was very little consensus between the draft ratings of the reviewers and the final rating that was assigned to the case study. In the example of Figure C.4, three of the six draft ratings (50%) exactly matched the final impact category rating. There was one occurrence of ±1 (17%) and two occurrences of ±3 (33%). For the purposes of our analysis, if the majority of individual draft ratings were either equal to or differed from the final rating by a single rating point, this indicated good levels of intra-panel reliability. As with most peer review processes, there are always issues of subjectivity associated with individual judgements; nevertheless, we remain relatively confident that the thresholds as we have set them are within the realms of 'normal' or 'expected' deviations and subtle disagreements in scoring preferences.

The results of the analysis of data from each of the panels are now presented in turn.

C.4.2 Economic Development Panel

There were two sub-panels within the Economic Development Panel; 13 case studies were submitted to sub-panel 1 and scoring templates from seven reviewers have been analysed in this section. Not every reviewer assessed all 13 case studies, resulting in an aggregate of 66 draft ratings. In sub-panel 2, there were 14 case studies in total (three of which were also assessed by sub-panel 1) and scoring templates have been analysed from six reviewers. There were 77 draft ratings received for sub-panel 2.

Intra-panel reliability

Combining data from both sub-panels, almost 85% of the draft ratings were either equal to their resepective final ratings or differed from the final ratings by a single rating point, demonstrating high levels of agreement between the individual reviewer ratings and the final ratings. The distributions of the deviations between the assessor draft ratings and the final ratings for the Economic Development Panel case studies are shown in Figure C.5 (the results have been presented for data from both sub-panels as well).

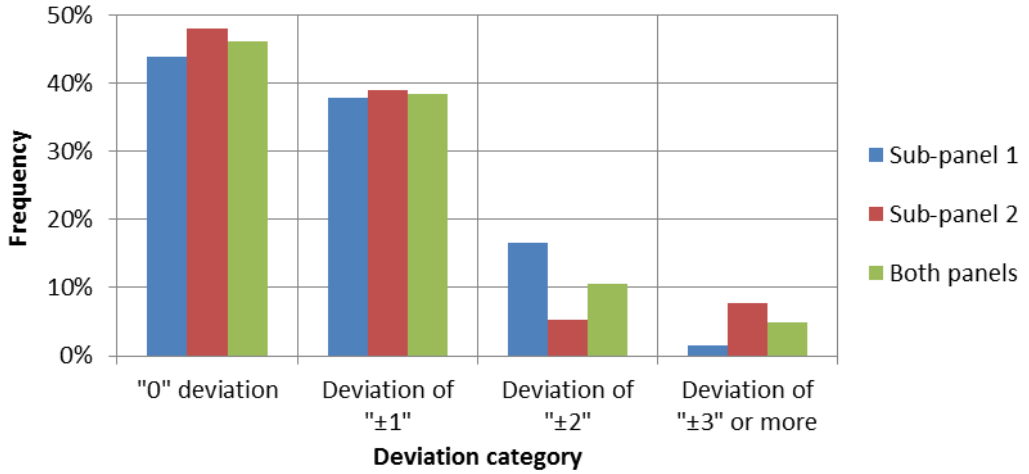

Figure C.5: Occurrence frequencies of instances where the individual draft ratings varied from the final ratings by zero, one, two and three or more rating points for case studies submitted to the Economic Development Panel

To illustrate the levels of consensus between the draft ratings and the final ratings for each of the reviewers in the Economic Development Panel, the occurrence frequencies (expressed as percentages) within each of the four deviation classes are presented in Figure C.6 for all the reviewers. Nine of the 13 reviewers had at least 40% of their individual ratings exactly matching the final rating, and Reviewer 1 in sub-panel 2 had as many as ten of 14 draft ratings (71%) identical to the final rating. Overall, 12 of the 13 reviewers had at least 75% of their draft ratings either equal to or differing from the final rating by a single point, again demonstrating the high levels of intra-panel reliability within the Economic Development Panel.

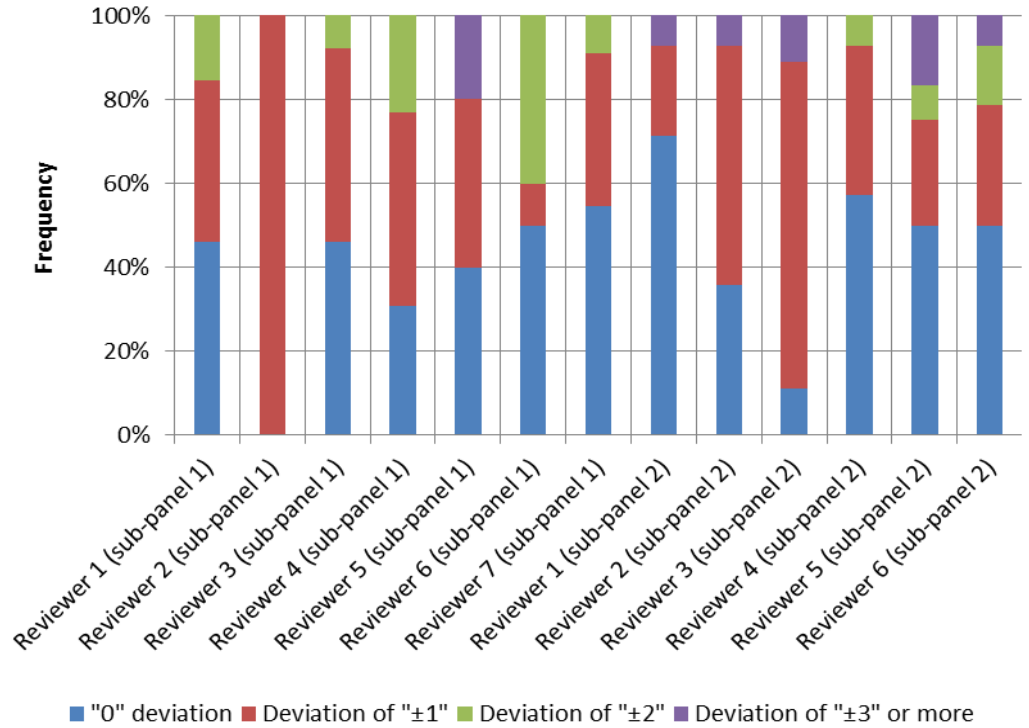

Figure C.6: The levels of consensus for each reviewer between the individual ratings and the final ratings for case studies submitted to the Economic Development Panel

Case study assessment times

To obtain an idea of the levels of burden involved in carrying out the impact case study assessments, the times that each of the reviewers took to assess their respective case studies were collated from the individual scoring sheets. For all the reviewers in the Economic Development Sub-Panel 1, the average time taken to complete the assessment of each of the case studies was 28.7 minutes (the most and least times spent on a case study within sub-panel 1 were 45 minutes and 20 minutes, respectively). The reviewers in sub-panel 2 took slightly less time (27.5 minutes) on average to complete each of their assessments (the most and least times spent on a case study within sub-panel 2 were 60 minutes and 11.5 minutes, respectively). The overall mean for both sub-panels was 28.0 minutes. The average assessment times per case study for all the reviewers in the Economic Development Panel have been listed in Table C.6.

Table C.6: The average time taken by reviewers on the Economic Development Panel to complete each case study assessment

Reviewer	Sub-panel	Number of case studies reviewed	Average time to complete assessment per case study (minutes)
Reviewer 1	1	13	30.0
Reviewer 2	1	1	25.0
Reviewer 3	1	13	30.0
Reviewer 4	1	13	23.8
Reviewer 5	1	5	38.0
Reviewer 6	1	10	28.3
Reviewer 7	1	11	29.1
Reviewer 1	2	14	17.6
Reviewer 2	2	14	20.4
Reviewer 3	2	9	60.0
Reviewer 4	2	14	17.1
Reviewer 5	2	12	37.5
Reviewer 6	2	14	23.9

C.4.3 Environment Panel

As with the Economic Development Panel, there were two sub-panels within the Environment Panel. Sub-panel 1 assessed six case studies, and scoring templates from five reviewers have been analysed. There was an aggregate of 28 draft ratings distributed across the five reviewers and six case studies that were examined for levels of consensus with the final ratings. With regard to sub-panel 2, there were nine case studies in all (two of which were also assessed by sub-panel 1) and scoring templates have been analysed from one reviewer, resulting in a total of nine draft ratings.

Intra-panel reliability

Combining data from both panels, 30 of the 37 draft ratings (approximately 81%) were either equal to or differed from the respective final ratings by a single rating point, once again signifying a high degree of consensus between the various reviewers. Almost 46% of the draft ratings were identical to the final ratings that were allocated to the corresponding case studies, indicative of

very good levels of intra-panel reliability. Of the four panels, the Environment Panel was the only one that did not have any individual assessor ratings that differed from the final ratings by three or more rating points, although compared with the other three panels, the Environment Panel had the maximum fraction of '±2' deviations (almost 19%). The distributions of the deviations between the draft ratings and the final ratings for the Environment Panel are illustrated in Figure C.7 (along with the combined sub-panel data, the data from both sub-panels have also been presented).

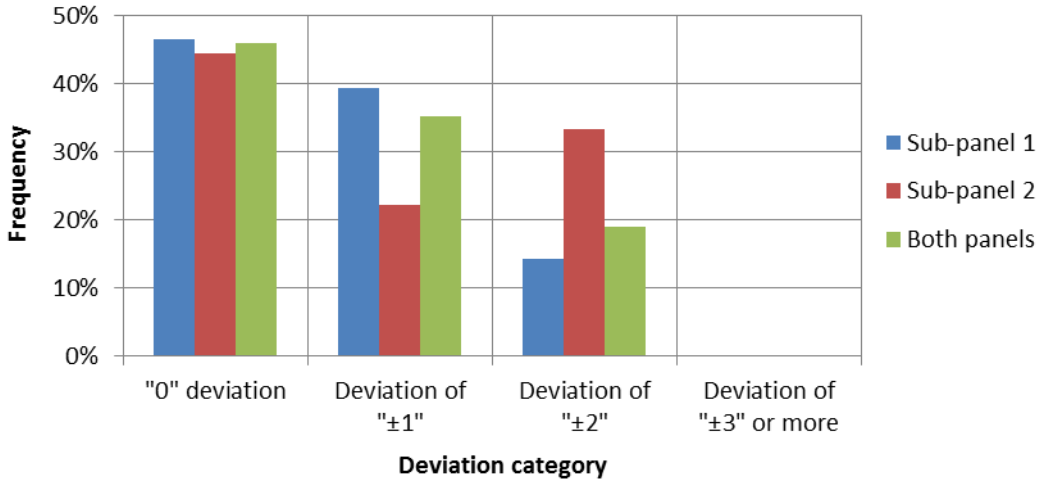

Figure C.7: Occurrence frequencies of instances where the individual draft ratings varied from the final ratings by zero, one, two and three or more rating points for case studies submitted to the Environment Panel

The occurrence frequencies (expressed as percentages) within each of the four deviation classes are presented in Figure C.8 for all the reviewers within the Environment Panel. All six reviewers had at least two-thirds of their individual draft ratings either identical to or differing from the final ratings by a single rating point (in sub-panel 1, the corresponding fraction was at least 80%), again illustrating good levels of intra-panel reliability.

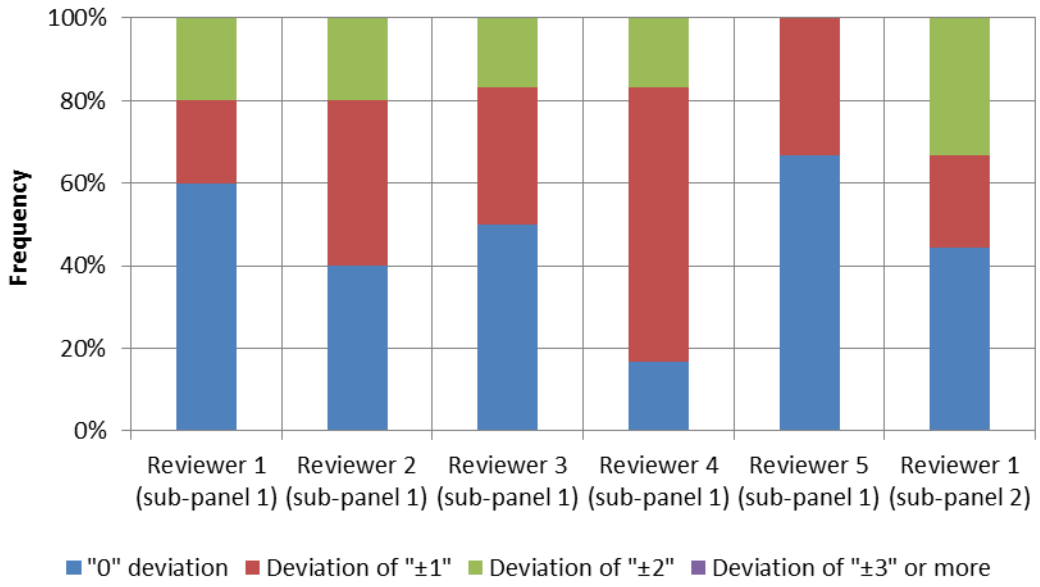

Figure C.8: The levels of consensus for each reviewer between the individual ratings and the final ratings for case studies submitted to the Environment Panel

Case study assessment times

The average time taken to assess each of the case studies for the reviewers in sub-panel 1 was 30.9 minutes (the most and least times spent on case studies within sub-panel 1 were 55 minutes and 17 minutes, respectively). The reviewer in sub-panel 2 took approximately 45 minutes on average to complete each assessment (the most and least times spent on case studies within sub-panel 2 were 75 minutes and 30 minutes, respectively). The overall mean for both sub-panels was 34.4 minutes. The average times taken by the Environment Panel reviewers to assess each of their case studies have been listed in Table C.7.

Table C.7: The average time taken by reviewers on the Environment Panel to complete each case study assessment

Reviewer	Sub-panel	Number of case studies reviewed	Average time to complete assessment per case study (minutes)
Reviewer 1	1	5	26.3
Reviewer 2	1	5	24.2
Reviewer 3	1	6	31.7
Reviewer 4	1	6	23.2
Reviewer 5	1	6	51.0
Reviewer 1	2	9	45.0

C.4.4 Society Panel

There were two sub-panels within the Society Panel. Sub-panel 1 assessed 12 case studies and scoring templates from three reviewers have been analysed. Because of the workload in the Society Panel, seven case studies were not given a final rating (they were 'unrated') even though they were assessed by the individual reviewers. Excluding these unrated case studies, there were a total of 14 draft ratings distributed across the three reviewers and the five case studies that were examined for levels of consensus with the final ratings. For sub-panel 2, there were 13 case studies (two of which were also assessed by sub-panel 1) and scoring templates have been analysed from six reviewers. One of the case studies was 'unrated' resulting in a total of 61 draft ratings.

Intra-panel reliability

Slightly more than 77% of the draft ratings were either equal to their resepective final ratings or differed from the final ratings by a single rating point, again highlighting good levels of intra-panel reliability. The distributions of the deviations between the draft ratings and the final ratings for the Society Panel are presented in Figure C.9 (the data from both sub-panels have been presented along with the combined data). Compared to the Economic Development (approximately 46%), Environment (approximately 46%) and Defence (approximately 44%; presented in C.4.5) panels, however, there was a significantly smaller proportion of cases (approximately 31%) where the draft ratings and the final ratings exactly matched. Indeed, the Society Panel was the only one in which the proportion of ±1 deviations exceeded the proportion of 0 deviations.

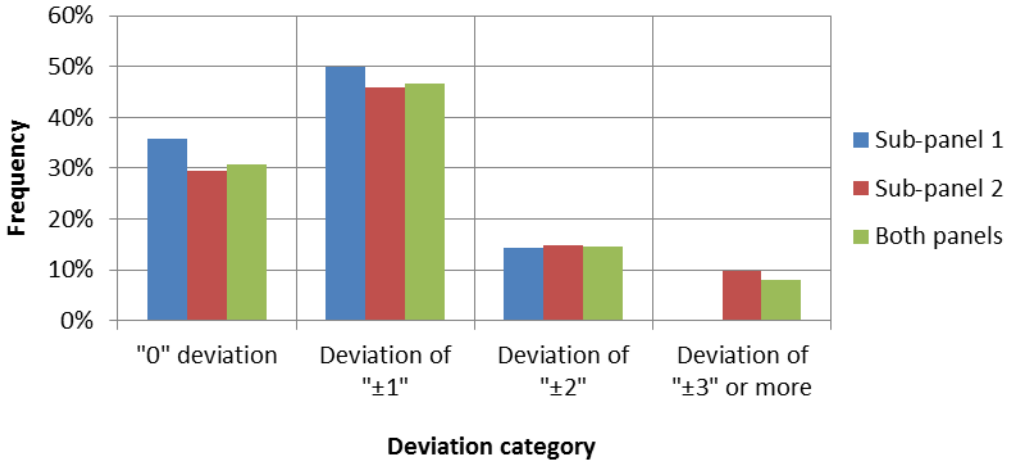

Figure C.9: Occurrence frequencies of instances where the individual draft ratings varied from the final ratings by zero, one, two and three or more rating points for case studies submitted to the Society Panel

To examine the levels of agreement between the draft ratings and the final ratings for each of the reviewers in the Society Panel, the occurrence frequencies (expressed as percentages) within each of the four deviation classes are presented for each reviewer in Figure C.10. Once more indicating relatively good levels of intra-panel reliability, seven of the nine reviewers had at least 20% of their individual ratings identical to the final ratings, and all nine reviewers had at least 58% of their draft ratings either equal to or varying from the final rating by a single rating point.

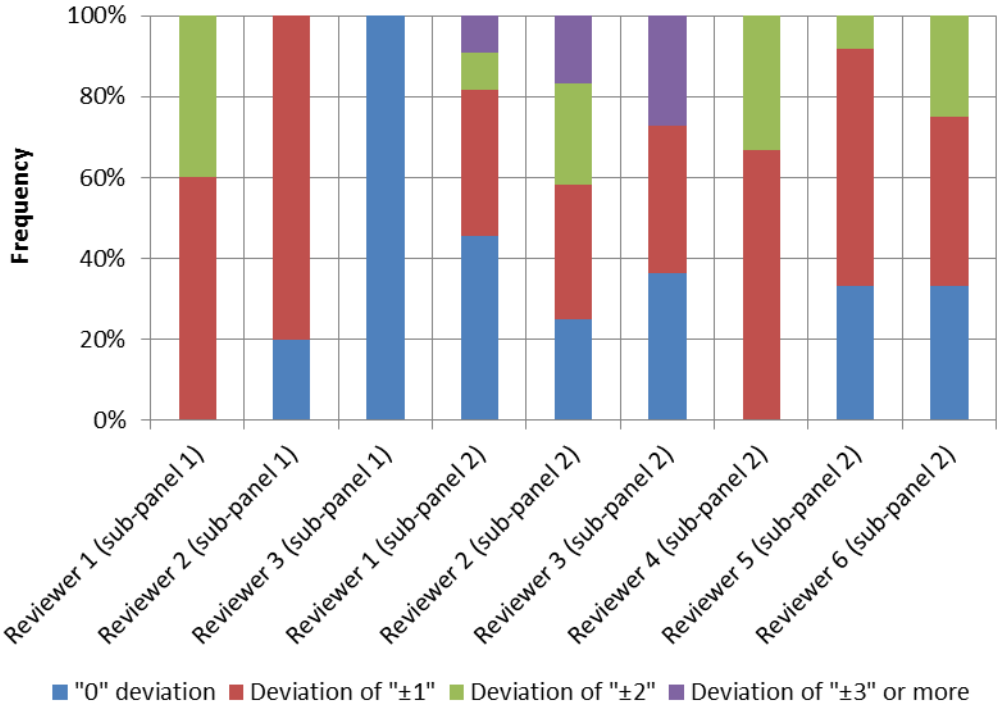

Figure C.10: The levels of consensus for each reviewer between the individual ratings and the final ratings for case studies submitted to the Society Panel

Case study assessment times

The average times taken by each reviewer to assess their respective case studies in Society sub-panels 1 and 2 have been listed in Table C.8 (the times spent by reviewers to assess case studies that were ultimately 'unrated' have also been included). The average assessment time per case study for the reviewers in sub-panel 1 was 30.4 minutes (the most and least times spent on a case study within sub-panel 1 were 45 minutes and 12 minutes, respectively). The reviewers in sub-panel 2 took approximately 25 minutes on average to complete each of their assessments (the most and least times spent on a case study within sub-panel 2 were 40 minutes and 15 minutes, respectively). The overall mean for both sub-panels was 27.0 minutes.

Table C.8: The average time taken by reviewers on the Society Panel to complete each case study assessment

Reviewer	Sub-panel	Number of case studies reviewed	Average time to complete assessment per case study (minutes)
Reviewer 1	1	12	27.1
Reviewer 2	1	12	30.4
Reviewer 3	1	1	35.5
Reviewer 1	2	12	20.0
Reviewer 2	2	13	33.5
Reviewer 3	2	12	21.7

Reviewer 4	2	4	30.0
Reviewer 5	2	13	25.0
Reviewer 6	2	13	19.5

C.4.5 Defence Panel

Four case studies were submitted to the Defence Panel. Five of the six reviewers assessed all four case studies while one reviewer assessed three of the case studies, resulting in a total of 23 draft ratings. Even though there were significantly fewer data in the Defence Panel than in the other panels, the results have been presented for the purpose of completeness.

Intra-panel reliability

Approximately 78% of the draft ratings were either equal to their resepective final ratings or differed from the final ratings by a single rating point, which indicates that overall there were good levels of consensus in the Defence Panel. The distribution of deviations for the Defence Panel case studies is shown in Figure C.11.

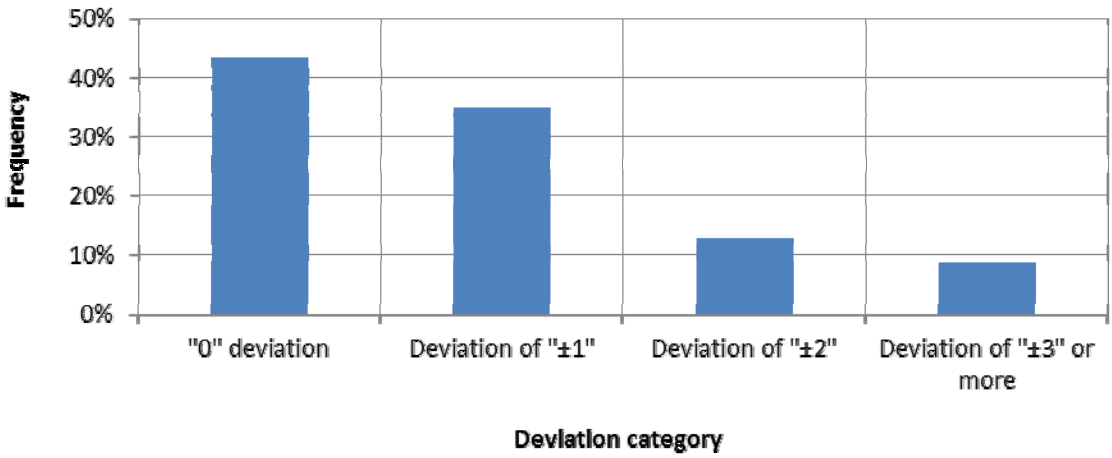

Figure C.11: Occurrence frequencies of instances where the individual draft ratings varied from the final ratings by zero, one, two and three or more rating points for case studies submitted to the Defence Panel

To gain an idea about the levels of consensus between the draft ratings and the final ratings for each of the individual reviewers in the Defence Panel, Figure C.12 displays the occurrence frequencies (expressed as percentages) within each of the four deviation classes for the six reviewers. The distribution of data in the figure shows that there were relatively high levels of consensus within the panel and relatively little movement after the panel meetings. Specifically, for reviewers 1, 2, 3 and 6, all their individual ratings either matched the respective final ratings or differed from the final ratings by a single rating point, indicating high levels of consensus. With particular reference to reviewers 1, 2 and 3, three of their four ratings were identical to the final ratings given to the corresponding case studies.

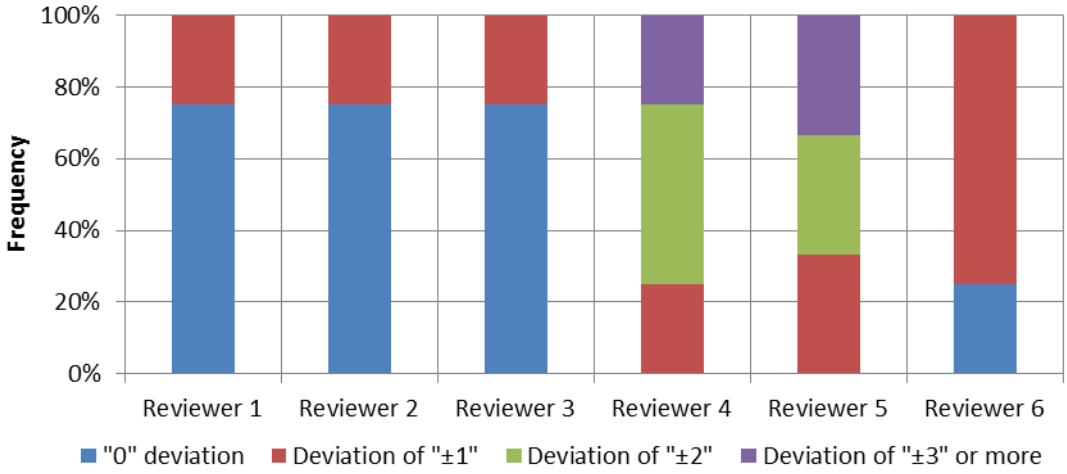

Figure C.12: The levels of consensus for each reviewer between the individual ratings and the final ratings for case studies submitted to the Defence Panel

Case study assessment times

For all the reviewers in the Defence Panel, the average time taken to complete the assessment of each of the case studies was 28.4 minutes (the most and least times spent on a case study within the panel were 40 minutes and 17 minutes, respectively). The average times taken by the reviewers to assess each of their case studies have been listed in Table C.9.

As mentioned previously, reviewers 1, 2 and 3 displayed the highest levels of consensus between their individual ratings and the final panel ratings. Table C.9 shows that reviewers 1 and 3, on average, took the least times to complete each of their assessments, and Reviewers 4 and 5 took the most times (on average) to assess each of their case studies, yet displayed the least levels of consensus between their respective draft ratings and the final ratings assigned to the case studies they reviewed. This highlights the issue of subjectivity associated with peer review processes, and that, at least in the case of the Defence Panel, there was no direct correlation between the levels of scoring consensus and burden involved in assessing the case studies (just because a reviewer spent less time assessing a case study did not mean that there were greater chances of them 'getting it wrong', and vice versa).

Table C.9: The average time taken by reviewers on the Defence Panel to complete each case study assessment

Reviewer	Number of case studies reviewed	Average time to complete assessment per case study (minutes)
Reviewer 1	4	19.8
Reviewer 2	4	31.3
Reviewer 3	4	22.5
Reviewer 4	4	36.3
Reviewer 5	3	33.3
Reviewer 6	4	28.8

C.4.6 All panels combined

Intra-panel reliability

The draft ratings from all four panels were combined and compared to the corresponding final ratings to obtain an overall picture of the levels of consensus between individual draft ratings and corresponding final ratings allocated to the ATN university case studies. Overall, almost 42% of the draft ratings exactly matched the final ratings, and approximately 40% of the draft ratings differed from the final ratings by a single rating point, which demonstrates that across the trial, in general, there was very good intra-panel reliability. The results have been presented in Figure C.13, and for comparison the distributions of deviations for the four panels have also been included. To recap, the approximate proportions of cases where the individual draft ratings were either identical to or varied from the final impact category ratings by a single point for the four expert panels were as follows: Economic Development: 85%, Environment: 81%, Society: 77%, and Defence: 78%. It is also worth noting that 97% of the assessor draft ratings used were between 'A' and 'D', suggesting that the reviewers may have collapsed a six scale into a four point scale. Alternatively they may have legitimately not seen the need to use 'E' or 'unclassified'.

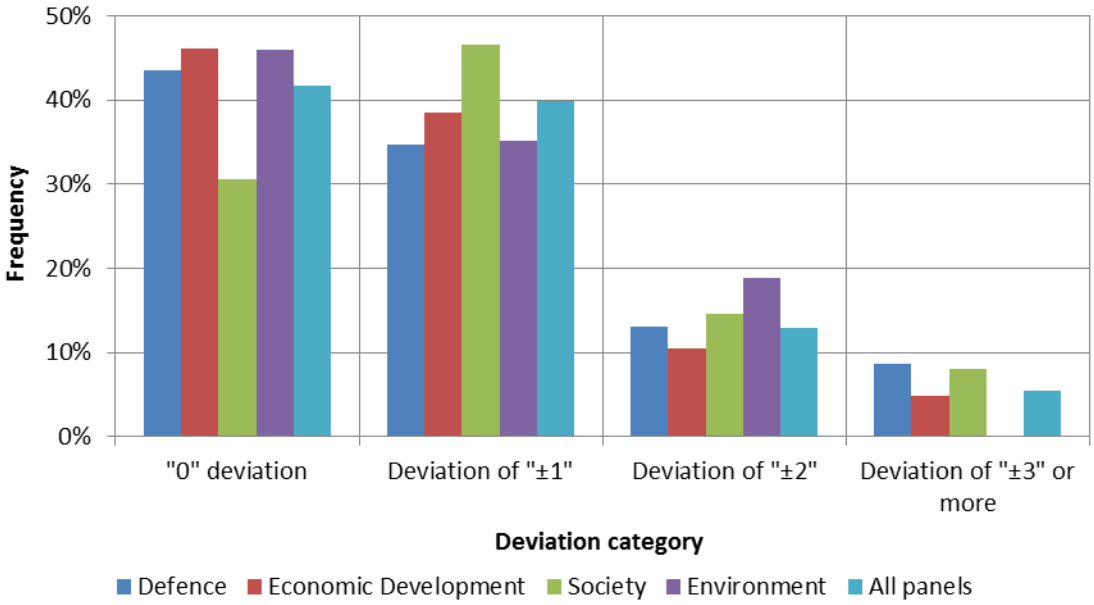

Figure C.13: Occurrence frequencies of instances where the individual draft ratings varied from the final ratings by zero, one, two and three or more rating points for all four panels, and when combining data from all the panels

Case study assessment times

Across all the panels, the average assessment time for each reviewer to assess a single case study was 28.4 minutes. For each reviewer within a particular panel, the average assessment time per case study was: 28.4 minutes (Defence), 28 minutes (Economic Development), 27 minutes (Society) and 34.4 minutes (Environment). Furthermore, the average total time spent on assessing each case study (across all the panels) was 112.1 minutes, with each case study being assessed by an average of four reviewers.

C. 5. Analysis of scoring data from sections 2, 3 and 4 of the individual scoring sheets

Finally, the scoring data pertaining to the following three sections in the individual scoring sheets have been analysed:

- 'rating of description of how the research underpinned the impact' (section 2 in the scoring sheet; variable name in the analysis: *R1*)
- 'rating of description of the nature and extent of the impact' (section 3 in the scoring sheet; variable name in the analysis: *R2*)
- 'rating of the validation of the impact' (section 4 in the scoring sheet; variable name in the analysis: *R3*).

Of all the data in the scoring sheet, it was felt that the scores provided in these three sections would potentially have had an influence in the way reviewers assigned draft ratings to case studies. In particular, the data were analysed to test if there is a statistically significant correlation between these three scores and the draft rating assigned to a particular case study. Much of the underpinning analysis carried out in this section was complex; this is a summary of the main results:

- The analysis of the data demonstrated that if a case study receives higher ratings in sections 2 (*R1*), 3 (*R2*) and 4 (*R3*) of the reviewer scoring sheet, it is likely to receive a higher draft rating.
- In particular, among the three scores, the score in section 3 (*R2*: 'rating of description of the nature and extent of the impact') had the maximum influence on the draft rating that was awarded to a case study.

The methodology employed to arrive at these results has been described below.

For the purpose of this analysis the letter scores of the assessor draft ratings were converted to numeric scores using the conversion scheme presented in Table C.1. However, since there were relatively few case studies that were scored 0, 1 or 2, it became necessary to merge categories of draft ratings and scores in sections 2, 3 and 4, according to the scheme presented in Table C.10.

Table C.10: Coding of data for regression analysis

Variable	Code
Draft rating	A (2), B(1), C or below (0)
R1	1 (high score 4 or 5), 0 (low score 1 to 3)
R2	1 (high score 4 or 5), 0 (low score 1 to 3)
R3	1 (high score 4 or 5), 0 (low score 1 to 3)

The 'recoding' of the data allows us to test if obtaining a high score in *R1*, *R2* and *R3* is correlated with the draft rating of a case study. Table C.11 presents the results of an ordered logistic regression[71] fitted to data from all the case study ratings. The modelling results indicate that there

[71] Further details about ordered logistic regression can be found in Torres-Reyna, O. (2013) *Getting Started in Logit and Ordered Logit Regression: Data and Statistical Services*, Princeton University.

is a statistically significant correlation between the draft rating and the scores *R1*, *R2* and *R3*.[72] Specifically, when *R1* changes from a lower score (1 to 3) to a higher score (4 or 5) the odds of getting a higher draft rating are 3.5 times greater. Similarly, higher scores for *R2* and *R3* increase the odds of getting a higher draft rating by 4.8 and 2.9 times, respectively. Thus, simply put, if a particular case study receives higher scores in sections 2, 3 and 4 of the scoring sheet, it is more likely to receive a higher draft rating, with the score *R2* ('rating of description of the nature and extent of the impact') having the most influence among the three scores.

Table C.11: Results of ordered logistic regression modelling

		Coefficient *B*	95% confidence interval for coefficients		Odds ratio *Exp(B)*	Significance
			Lower bound	Upper bound		
Threshold	[DR = .00]	2.539	1.852	3.225		.000
	[DR = 1.00]	4.617	3.815	5.419		.000
Location	*R1*	1.244	.565	1.923	3.5	.000
	R2	1.563	.838	2.287	4.8	.000
	R3	1.061	.446	1.675	2.9	.001

[72] Since the scales in *R1*, *R2* and *R3* were converted to binary (0, 1) variables, there was no specific adjustment related to the ordinal nature of these scales. It should be noted that we tested these scales before recoding them into binary (as the original five point scale) and the results were not very different.

Annex D : Survey results

D. 1. Methodology

A survey was developed to capture information about the experiences of institutions and case study authors who participated in the EIA trial. The questions to the case study authors focused specifically on the case study writing process (Figure D.1) and included topics such as time taken to complete the case study, rewards from the process, challenges and areas for improvement. The institutional questionnaire covered five main areas about the trial as a whole, including background and institutional strategy, interpretation and use of the guidance, approach and views on the case study template, institutional management of the case study process, and any indication of changing attitudes (Figure D.2). These figures are at the end of this annex.

The survey was circulated to participating institutions, where it was cascaded to the case study authors. Institutional responses[73] were received from all of the five ATN universities that participated in the trial (institutions A, B, C, D and E). From the 64 case studies submitted between the five universities in the ATN for the EIA trial, responses were received from 24 case study authors[74] (Table D.1).

Table D.1: Number of survey responses received from case study authors at each institution

Institution	Case studies submitted	Survey responses
Institution A	12	4
Institution B	14	4
Institution C	14	9
Institution D	11	4
Institution E	13	3

Analysis of the survey data proceeded in a grounded, iterative fashion, with thematic areas drawn out across the institutional and case study author responses. These were then reviewed in light of the wider headlines cutting across the review and the themes were then refined or new hypotheses posed about the data. This process was repeated until the themes described below were identified.

[73] These were completed by the director of research or delegated to the appropriate individual within the institution.

[74] The sample is biased towards Institution C as 37.5% of the responses were from this university. However, across the sample appeared to be representative of the range of ratings achieved and panels submitted to.

D. 2. Institutional responses

D.2.1 Guidance

Unit of analysis

Two of the five institutions specifically stated that SEO codes are appropriate units of analysis for the EIA Trial, as opposed to disciplinary units of analysis, and another commented that the analysis groupings were acceptable. However, three of the five institutions explicitly stated that the four primary sectoral divisions under which case studies must be submitted are unhelpful. More specifically, there was wide consensus among the three institutions that the society division should be broken down into sub-divisions. Two institutions offered various potential sub-categories including health; education, law and politics; cultural understanding; business; and creative industries. There was concern in one case that the defence division is too narrow. One institution also noted that in order to select the most relevant four-digit SEO codes, they needed to 'drill down' to the six-digit SEO code level and warned that there may be difficulties comparing 'apples to apples' in the assessment panels, if others do not disaggregate to the six-digit SEO code level. Another institution suggested that there should be guidance on how case studies should be approached when they apply to several of the four primary categories. This is an important consideration for improving the assessment of impact through a process like the EIA.

Examples of impact

All institutions generally agreed that the definition of impact in the guidance was clear and well defined, but three commented that best-practice examples of impact and case studies would be useful in understanding what should be included in each case study – and how impact should be approached. Therefore, it may be beneficial to publish a selection of case studies that demonstrate a range of impacts including the highest-scoring impact case studies in the EIA trial before repeating the exercise.

Quality of research

The guidance states that impacts should be based on quality research, although there is no quality threshold. Two institutions suggested that this should be clarified in the guidance – and one specifically noted that it would be useful to know what proportion of a case study should be committed to the relative research excellence of a case. A third institution noted that a 'suite of metrics' would be useful to aid evaluation in representing 'quality research'.

Timelines

Four of the five institutions agreed that the notice period was particularly short for the EIA trial; one believed that further iterations of the process would take less time, given that data collection will be incorporated within standard processes and procedures in the future, and another felt that the trial timelines were fine.

D.2.2 Case study template

Impact case study structure

Three of the five surveyed institutions noted that the structure of the impact case study could be improved. Two stated that the case study template should be re-ordered and another noted the necessity for repetition within the current template. One specifically stated that section 7 concerning underpinning research should precede section 6 on details of the impact – bringing

the template in line with the one used in the UK REF exercise. Concerns about repetition within the template implies that clarifying exactly what should be included under each section, along with examples, would be beneficial. This is explored further in relation to the case study summary and context below.

One institution noted that the 'template was satisfactory although the main suggestion would be to present it electronically through a web based portal'. This may improve analysis and reduce the institutional burden of the process.

Case study summary and context

Two of the institutions demonstrated confusion regarding the case study summary and context section requirements. One commented that there is overlap between the two sections, and suggested that they should be merged. This reveals a clear misunderstanding of what should be included in each section (and indeed this appeared to be the case within many of the reviewed case studies – see Annex B).

Another institution commented that the ability to refer to external web-links without further clarification of how this should be approached was unhelpful – and was exacerbated by the failure to enforce word limits (although it is worthy of note that another institution found the word limits helpful). More specifically, the lack of understanding about how much information should be included in each section and through external links made it difficult to know how best to demonstrate the impact. Similarly, another institution noted that 'it is unreasonable to assume that panel members will have the time and capacity to access material via the web or other sources'. This raised concerns about the readability and presentation of the case study. However, it is important to note that the guidance does not suggest the use of external links within these sections, and they did not appear in many of the reviewed case studies. Perhaps this simply requires a point of clarification within the guidance for each section.

D.2.3 Case study process

Engagement with end users

When asked how impact was identified within the research institution's portfolio, only one of the institutions outlined engagement with end users or stakeholders – which took the form of 'company feedback on use'. Instead methods for identifying impact included assessment of internal performance data, feedback from external partners and reviewers, economic assessments, roundtable discussions with staff, and consulting senior researchers and senior research administrators who are most familiar with the research. One institution conducted a web-based survey, but it is unclear who the survey was targeted at, and only two of the five institutions stated that end-user engagement had been sought to develop some case study drafts. This suggests there was a significant lack of end-user engagement before and during the trial, which should be improved before the full evaluation exercise.

Administrative burden

All of the surveyed institutions prepared more case studies than they eventually submitted – usually in order to make the final selection process. Three of the five institutions created formal steering or working groups to approve case studies for trial assessment, and in one case to coordinate and develop case studies as well. Institutions provided from one to three authors or researchers to develop the case study. Generally two support staff worked on each case study,

although one institution reported there was an average of three support staff on each case study (see Table D.2). Typically, research staff wrote and developed the document, with support staff reviewing and providing strategic and technical assistance. Unsurprisingly, then, all institutions reported a high level of administrative burden on staff time, particularly as mechanisms are not currently in place to capture impact systematically. One expressed concern that the employment of external assistance in order to aid with case study development may set a precedent for all institutions – which may then become a burden on the sector. However, another institution noted that the time burden would be justified if funding is a potential outcome.

Table D.2: Average number of research and support staff employed on a case study per institution

Institution	Average number of researchers per case study	Average number of support staff per case study
Institution A	1.7	1.7
Institution B	2–3	2
Institution C	1	2
Institution D	1	2
Institution E	1	3

Key challenges

Certain challenges were repeated several times across numerous institutions. These included identifying research with 'appropriate' impact or 'real examples' of impact and gathering verifiable evidence (particularly for older research); engaging the academic community to participate post ERA submission; and ensuring consistency and quality across all case studies in time for the deadline. Other challenges included lack of time, lack of data, availability of researchers associated with impact, lack of corporate knowledge and developing case studies with cultural rather than economic benefits. All three of the key challenges (capturing and identifying impact, engaging the academic community and ensuring consistency and quality) should be less problematic as the exercise is repeated and impact capturing mechanisms become embedded within the institution.

D.2.4 Changing attitudes

Mechanisms to capture impact

Three institutions specifically commented that the trial has highlighted the need to introduce mechanisms to capture impact systematically and regularly, particularly given that several case study authors were not familiar with the impacts of their research. One institution noted that it has already set up such a mechanism.

Marketing

All the institutions commented that the exercise will aid in producing marketing materials for the university, although one commented that the exercise revealed the university's existing strengths in publicising research impact.

D. 3. Case study author responses

D.3.1 Case study format

Just over three-quarters – 19 out of 24 authors (76%) – of those who responded to the survey felt that the format of the case study was generally appropriate to describe the types of impact their research had had. However, several authors mentioned they had difficulties in breaking the information down under different headings. They felt that it was unclear what belonged in each section and this led to repetition, especially between the impact and underpinning research sections. This was exacerbated by the order of the sections where the details of impact were required before the underpinning research, so the research was described twice to enable the reader to contextualise the impact.

D.3.2 Time taken

The length of time to complete a case study varied greatly, with case study authors estimating it took between two hours and three weeks. Some describe it as an iterative process occurring over a number of weeks, whereas others appear to have completed it in a short number of hours. To compare these, the number of hours spent was calculated, working on the assumption that there were eight hours in a work day. The average length of time spent on writing a case study was approximately 5 days and 6 hours (45 hours), although this figure was increased by some, which took several weeks; perhaps a more reflective figure is the median, which was 3 days and 2 hours (26 hours).

D. 4. Rewards and challenges

The authors were asked to identify rewards and challenges of participating in the assessment, and areas where improvements could be made for future assessments. The main rewards identified were recognition of the value of research, the ability to adapt materials for other purposes, understanding the impact and having the ability to capture the impact of their research, appreciating the need to collect evidence was ongoing, and consolidating networks. Challenges and areas for improvement included time taken, level of feedback, guidance and support, and challenges in identifying impacts. These are discussed below.

D.4.1 Reward: recognition of the value of research

Seven case study authors (29%) said that a reward of the process was the recognition of the value of their research received from others, especially the university, funders and those in the wider society. Five specifically commented on the value perceived at a university or faculty level and the reward for having represented their university in this process and disseminated information about their research and its benefits.

D.4.2 Reward: being able to adapt the material for other purposes

Six case study authors (24%) found one of the most rewarding things about producing a case study was that it would be able to be used for other purposes, such as marketing, grant

applications and promotion. Because the text was written for a non-academic audience, they were able to use a more accessible style of writing to demonstrate the impact of their research.

D.4.3 Reward: the ability to capture and document the impact that research had had

Twelve case study authors (48%) noted that documenting the impact and benefit their research had led to was a reward of the assessment. They thought it was rewarding to see the benefit of research beyond the university and to see these impacts collated in a single document. One author mentioned the value in bringing together components from a broad research programme in a single narrative.

D.4.4 Reward: appreciating the need to collect evidence was ongoing

Four case study authors (16%) had found capturing and collating impact as an ongoing process was a key benefit of the trial. One stated that it had 'prompted me to ensure I collect evidence of this kind as a matter of course for all research projects I undertake'. Another described a reward of the assessment as the generation of 'a single repository of all the details, outputs and benefits of the research project for future use in a single place'. There were six authors who spent less than one day (eight hours) on their case study. Four authors had built on previously collected materials. The author of another case study (which took over a day to produce) also mentioned basing the case study on a previously commissioned report. This indicates that the burden is reduced if data are collected throughout the research project and researchers think about impact throughout the research, rather than solely at the end. One author summed this up by stating that the assessment had 'encouraged us to think about measuring impact and more generally, prepare for it in the future, and formally insert it into our strategic plan'.

D.4.5 Reward: understanding the wider context and impact of research

Eight case study authors (32%) noted that a reward in participating in the trial was translating their research into impact, understanding the impact of their research and being able to evaluate and articulate it. Authors stated that the assessment had improved their ability to articulate impact, to think about how research is translated, and where this process can be improved, and to identify and reflect on the impact of their work.

D.4.6 Reward: networking

Two case study authors (8%) had used the assessment to contact people who they had interacted with in the past to provide input. One author said this contact with a previous client may lead to future interactions.

D.4.7 Challenge: burden

Ten case study authors (40%) said the time taken was a challenge with the process, and two found that obtaining resources was also a challenge. Two authors thought that extending the timeline would improve the process, allowing for a more detailed review and strengthened quality and succinctness. One author noted the 'lack of opportunity to revise and clarify the case study'

due to the short lead time. Another suggested that the system would be improved if there was real cost recovery for the time spent by staff and external consultants engaged in producing the case studies.

D.4.8 Challenge: word limits

Three case study authors said that one of the challenges of the assessment was compressing their responses within the prescribed word limit.

D.4.9 Challenge: feedback

Five case study authors (20%) specified the need for feedback during the process and after scoring the submission. Two authors said they had spent a lot of effort on their submissions and would have appreciated detailed feedback to enable them to improve their case studies in the future. This would increase the time taken and burden on the panel, but the EIA may want to consider if it would be possible as a mechanism to engage with the research community to improve impact and ensure it is applied throughout research. Perhaps in future assessments, exemplar case studies could be provided to guide researchers creating impact case studies. Two authors suggested that example case studies would be a useful resource.

D.4.10 Challenge: guidance and support

Seven case study authors stated that more guidance and support was required. One respondent defined the challenge as 'understanding clearly what was required for the case study', and another 'working back from the impact to how our research created it'. Another thought there was a need for further support for those involved in the process, possibly in the form of training in preparation for the assessment to understand what impact is and what a good case study looks like. In addition, universities need to support researchers during the process in order to identify and articulate impact. For example, one author stated that they had difficulty in creating the narrative of the case study and two suggested that engaging professional non-academic writers to support the researchers in development and crafting the case study would improve the system.

D.4.11 Challenge: developing evidence

Thirteen case study authors reported difficulties with collecting evidence. Three had difficulties in trying to define what could be claimed and quantified. Four noted that this required collaboration with other academics and end users of the research to develop the full narrative, which could be challenging and involve issues such as confidentiality and tracing people from the past.

Potential improvements to developing evidence include commencing collating supporting data earlier and providing clear definitions of what would be accepted as evidence of impact. One author thought there should be a panel-specific view on the potential weighting for different types of measures to define which measures and outcomes were more important in different sectors.

D.4.12 Challenge: template order

Three case study authors commented that the order of the template should be changed so that the underpinning research (currently question 7) preceded the impact (currently question 6). It was also felt by two respondents that there was overlap in the sections and one suggested a series of more intuitive subheadings that they felt would be useful to guide the writer.

D. 5. Survey questions

The survey question for case study authors and institutions are below.

DISCLAIMER: The information obtained through this document will only be used in the independent process evaluation of the EIA trial and will be anonymised and treated as confidential

EIA process evaluation survey for case study authors

1. What was the name of your case study? (Please note all comments will be anonymised. We are only asking this to enable analysis of separate disciplines)

2. How long did it take you to produce the case study?

3. How appropriate was the case study format to describing the different types of impact that you identified?

4. What were the three most rewarding things about developing your case study?

5. What were the three most challenging things about developing your case study?

6. What are the three things you think could be improved about the process?

Figure D.1: Survey questions for case study authors

DISCLAIMER: The information obtained through this document will only be used in the independent process evaluation of the EIA trial and will be anonymised and treated as confidential

EIA process evaluation survey for institutions

Background

1. What did you want to achieve by your participation in the Excellence in Innovation for Australia (EIA) trial?

2. What did you achieve through your participation of the EIA trial?

3. What was your institutional strategy for identifying and selecting your case studies?

Section 1 – Guidance

4. How satisfied were you with the guidance provided and what are the three main recommendations would you make to improve the guidance?

5. How clear was the definition of research provided in the guidance?

6. How clear was the definition of impact provided in the guidance?

7. How relevant were the disciplines within which topics were clustered for the evaluation of impact? Would you suggest a different grouping?

8. Were the guidelines sufficiently clear in regards to what counted as impact and accurate in relation to what was required of case study authors and the institution?

9. How realistic was the timeframe for producing your case studies?

Section 2 – Case study template

10. How satisfied were you with the template provided and what are the three main recommendations would you make to improve it?

11. Did you feel like the format allowed you to adequately capture impacts and articulate everything which you needed to share to showcase that impact?

Section 3 – Case study process

12. How did you identify impact within your institution's research portfolio?

13. Did you prepare more case studies than you eventually submitted, and if so how did you shortlist the ones which made up your final submission?

14. Who coordinated, developed and approved each individual submission?

8. Were the guidelines sufficiently clear in regards to what counted as impact and accurate in relation to what was required of case study authors and the institution?

9. How realistic was the timeframe for producing your case studies?

Section 2 – Case study template

10. How satisfied were you with the template provided and what are the three main recommendations would you make to improve it?

11. Did you feel like the format allowed you to adequately capture impacts and articulate everything which you needed to share to showcase that impact?

Section 3 – Case study process

12. How did you identify impact within your institution's research portfolio?

13. Did you prepare more case studies than you eventually submitted, and if so how did you shortlist the ones which made up your final submission?

14. Who coordinated, developed and approved each individual submission?

Section 4 - Changing attitudes

22. Were there any new insights about impact or opportunities to pursue further impact which arose out of the submission process?

23. How will you be using the results of the case study review to inform activities within your institution? We are interested in both internal activities, i.e. systems set up to capture impact for future submissions, and external activities, such as development of promotional material.

24. Did participation in the trial contribute to any changed attitudes of researchers and, if so, how?

Section 5 – Closing questions

25. What are the three things you think could be improved about the entire process?

26. Do you have any other comments/concerns you would like to share?

Figure D.2: Survey questions for institutions