THE AUSTRALIAN NATIONAL UNIVERSITY

Research Evaluation and Policy Project

Research School of Social Sciences

Testing Quantitative Indicators of the Quality and Impact of Research in the Social Sciences: A Pilot Study in Economics

Claire Donovan and Linda Butler

REPP Discussion Paper
05/2
September 2005

Donovan, Claire and Linda Butler
Testing Quantitative Indicators of the Quality and Impact of Research in the Social Sciences: A Pilot Study in Economics
E-mail: claire.donovan@anu.edu.au
linda.butler@anu.edu.au

REPP Discussion Papers are scholarly papers that report research in progress. They can be downloaded free of charge (PDF)

Research Evaluation and Policy Project
Research School of Social Sciences
The Australian National University
Canberra ACT 0200
Tel: (02) 6125 4849
Fax: (02) 6125 9767
http://repp.anu.edu.au

CONTENTS

TABLES

1. INTRODUCTION

1.1 Context and purpose of the pilot study

As we write this paper, the Department of Education, Science and Training (DEST) is deliberating adopting a Research Quality Framework (RQF), premised upon assessing the quality and impact of publicly-funded research, as the basis for distributing block funding for university research. In the RQF context, 'quality' is defined as the intrinsic merit of research and its influence upon academia, and 'impact' is defined as the effects of research beyond academia (i.e. its social, cultural, economic and environmental usefulness). It has been signalled that the RQF should be underpinned by 'metrics', i.e. quantitative indicators of research quality and research impact, which are appropriate to the field of study being assessed and which can be internationally benchmarked. However, existing Australian and overseas attempts at research assessment rarely venture beyond using standard publication and citation-based indicators to quantify research quality (as defined by the RQF) which are problematic when applied to many fields of study,[1] and quantitative measures of research impact are mooted but have yet to be operationalised. The RQF is, then, advocating the use of quality and impact metrics although little is known about their practical application: a cause for concern in itself and even more so when these are to be linked to the distribution of funds. This paper is a first step towards addressing this knowledge gap, and reports the process and findings of a Research Evaluation and Policy Project (REPP) pilot study which sought to investigate on a comparative basis the viability of using a range of novel quantitative indicators as proxies for research quality and research impact, and whether these may be used with confidence to inform the process of distributing research funds.

This pilot study does not investigate the qualitative assessment of research quality or impact, which to date is not on the RQF agenda, but acknowledges that any future steps should embrace this approach as the distinctive qualities of various strands of publicly-funded research, and their value within and beyond academia, may be lost when dissected into numerical values.

This pilot study takes the view that no indicator should be applied to research assessment in isolation, and nor should any indicator or suite of indicators be used independently from discipline-specific expert peer consultation.

1.2 Acknowledgements

This pilot study grows out of a REPP Australian Research Council (ARC) Linkage Project 'Strategic Assessment of Research Performance Indicators', undertaken in conjunction with DEST and the University of Queensland, which focused purely on the quantitative assessment of research quality.[2]

We wish to thank the four economics units and one multidisciplinary research unit which agreed to participate in this pilot exercise. We would particularly like to express our gratitude to the members of the study's Reference Group for their time and effort devoted to providing

[1] We shall see that standard bibliometric indicators were originally developed to be applied to the experimental sciences, and are not well suited to the humanities, arts and social sciences. Indeed, these publication-based measures even present problems when applied to the science-based disciplines of computing and engineering.

[2] See http://repp.anu.edu.au/repp.htm for an outline of this project and details of a DEST workshop devoted to the research findings.

peer evaluation of quality and impact indicators which were ‘road-tested’ in a very tight timescale (the whole pilot study was conducted within seven weeks). Thanks are also due to Anne Hill, Research Assistant in REPP, for her research support and dedication to this task.

2 WHY ECONOMICS?

The choice of economics as the focus of this pilot study was partly pragmatic, yet underpinned by sound reasoning. Bibliometric analyses are most suited to the experimental sciences, with the creative arts and humanities being least well-served and hence more likely to benefit from the use of novel indicators. Social sciences occupy a middle-ground within this spectrum; economics (along with applied psychology) is relatively well-served by standard bibliometrics due to its highly quantitative nature and associated publishing norms, so it is a safe choice to take this discipline and explore whether non-standard bibliometric and other non-bibliometric indicators can tell us anything new about research quality and research impact when compared with standard approaches. Our task was also made somewhat easier by the fact that economists are fascinated with ranking economists, and several relevant rankings had already been produced within the discipline. For the comparative element of this pilot study, we could readily identify economics groups or departments (referred to as ‘units’ in this study) with quite different research interests and varying degrees of theoretical, applied and policy-oriented focus. Economists are also a natural choice to invite to consider data relating to, and issues surrounding, the testing and implementation of various quantitative indicators.

3 THE INDICATORS

This study taps into ideas that are very much in the air, but about which little is truly known or has been tested. This pilot research concentrates on novel quantitative indicators of research quality in the form of non-standard bibliometric measures and non-bibliometric (or 'esteem') measures, and quantitative indicators of research impact. The selected indicators are cherry-picked for their promise from a catalogue of 75 novel measures identified by REPP research into Australian and international indicators of research quality and impact.[3] These indicators are either currently used in primitive form, or are being proposed (but are yet to be operationalised) by research funding agencies or for national research evaluation exercises. All are attracting interest from disciplines where traditional publication-based or citation-based measures are difficult to apply. However, these indicators are yet to be tested in a rigorous, quantifiable manner, so their actual value remains uncertain.

These measures are supplemented by non-standard bibliometric indicators developed by REPP's ARC Linkage Project 'Strategic Assessment of Research Performance Indicators'.[4]

3.1 Quality Indicators

There are two main types of research quality indicators: bibliometric indicators (publication and citation-based) and non-bibliometric or 'esteem' indicators (based on the regard for individuals or pieces of research). This pilot study focuses on novel indicators, so in terms of bibliometric measures it eschews developing the already well-trodden ground of standard measures (based on publication counts or citations to ISI-indexed journal papers) and concentrates on two non-standard bibliometric indicators, although some standard measures are presented for comparison. The indicators chosen for testing are introduced below, together with the reasons for their selection.

3.1.1 Standard bibliometric indicators

While noting that this pilot study was to focus on novel indicators, we believed that presenting data for some standard bibliometric indicators would provide useful reference points. We selected three indicators that are routinely used in research evaluation, both by professional practitioners and 'amateur bibliometricians'. They can be constructed from readily available sources, such as universities' returns to DEST, or from ISI databases or products to which most researchers have access – the Web of Science and Journal Citation Reports. They are measures that are familiar to most researchers.

Number of DEST publications

Each year, every Australian university submits data to DEST detailing the number of research outputs published in the preceding year. The data is used as one of three elements in formulas that are currently used to distribute block funding for research to the institutions. The term 'DEST publication' has now entered into the vernacular of most Australian academics. Many

[3] Claire Donovan (forthcoming 2005) 'Setting the Scene: A Review of Current Australian and International Practice in Measuring the Quality and Impact of Publicly-Funded HASS Research', part of a Council for Humanities, Arts and Social Sciences (CHASS) report 'Measuring the Quality and Impact of Publicly-Funded HASS Research'.

[4] See REPP (2005) *Quantitative Indicators for Research Assessment – A Literature Review*, REPP Working Paper 05/1. http://repp.anu.edu.au/Literature%20Review3.pdf.

commentators have raised questions about the equitability of the current distribution of research funds in Australia, particularly in relation to the publications component, so this is an ideal measure to incorporate into our pilot study.

Citation per publication rates (ISI journals only)

Simple citation counts are routinely constructed in the assessment of research. Those constructing rankings tables of institutions or departments routinely incorporate this calculation. Many of ISI's own analytical products, such as Essential Science Indicators or University Science Indicators, are centred around this measure. Many researchers now produce such calculations when applying for positions or promotions. Its use in social science disciplines is less widespread than in the experimental sciences, but appears to be rapidly gaining pace despite concerns about the validity of applying this measure to the social sciences (see Sections 2 and 3.1.2).

Journal publications weighted by journal impact factor (ISI journals only)

Where actual citation counts for a publication are not known, a common surrogate is used to weight the counts by ISI's journal impact factor. This is popular where assessors do not have direct access to ISI data, or where the level of aggregation is large enough to make the use of the *Web of Knowledge* (and the limitations on extraction capabilities inherent in its design) impractical. It is now commonplace to see the ISI journal impact factor noted in CVs and on grant applications, particularly in the natural sciences.

3.1.2 Non-standard bibliometric indicators

Any standard bibliometric measure that is based on either citation counts or journal impact factors presents problems for the social sciences. It will relate to only a small fraction of the publication output of the unit – that portion appearing in journals indexed by ISI. In some disciplines this can represent less than 10 percent of total output, and rarely climbs above 30 percent. The two non-standard bibliometric indicators included in this study seek to achieve a much wider coverage.

Citations to all publications (i.e. not just ISI journal articles)

Standard citation indicators relate only to publications appearing in journals indexed by ISI. This provides very good coverage for most of the experimental sciences, but relatively limited coverage for the applied sciences and social sciences, and very limited coverage in the humanities. As part of its ARC Linkage Project, REPP and its collaborators at the University of Leiden sought to develop novel citation indicators that utilised ISI data on 'non-source' publications,[5] thus expanding the usefulness of these indexes for disciplines not well served by traditional measures. This pilot study provided the ideal opportunity to test one of newly developed measures – citation per publication rates for all publications.

Publication counts weighted by prestige

An indicator commonly proposed, particularly in the social sciences and humanities, is one which weights books and books chapters according to the prestige of the publisher; or journal articles by the prestige of the journals in which they are published.

[5] While ISI only unifies citation counts for articles appearing in the journals it indexes, its database does contain details of citations to other publications that appear in its journals. These are termed 'non-source' publications, and may be books, book chapters, conference papers, articles in non-ISI journals, or any other form of publication.

Testing the second indicator is particularly timely as there are several European initiatives currently underway to produce publication rankings based on the perceived prestige of journals within the field or according to the reputation of publishing houses. In 2005 an attempt at ranking (rather than banding) journals by the UK Arts and Humanities Research Council was abandoned in the face of resistance from the academic community who objected on a variety of grounds to the AHRC generating lists of the top ten 'most significant and important' journals in ten subject areas.

Working paper views and downloads (EconLit)

Accessed through the EconLit website, LogEc data on abstract viewing and document downloads are taken to denote the academic popularity of particular pieces of research. The LogEc registered author search function calculates the frequency of working paper and journal paper views and downloads for registered individuals over a twelve month period. This allows us to assess the significance to the discipline of non-standard publications, i.e. working papers.

3.1.3 Non-bibliometric indicators

Non-bibliometric – or what are commonly known as 'esteem' indicators – of research quality are measures which are not based on the number published research 'outputs' or citation rates, but on the standing of an individual or of pieces of research within the academic community. While these types of indicators are assumed to be important for areas where bibliometric indicators are difficult to apply, and in principle are thought to be equitable for all fields of study, they have rarely been tested so their value remains uncertain. All data has been extracted from the annual reports of the study's participating units.

Honours, awards and prizes

This non-bibliometric indicator is a simple count of the honours, awards and prizes received by individuals. These are taken to reflect the esteem or standing of academics as judged by their peers (and by extension the quality of the person's research). Aggregated, this indicator is taken to denote the relative standing of a unit or department.

Within Australia, AIMS, ANSTO and CSIRO list the awards and prizes received by their staff to denote research quality; ARC end of grant reports note national and international prizes, honours and awards to demonstrate the quality of published writing or an individual's research standing; and in 2005 the NHMRC is introducing measures for the performance indicator of 'Increased capacity of Australian researchers' including awards, prizes and other honours. In all cases this data is collected but not tied to the allocation of research funding. Internationally, the Netherlands national assessment of research records honours, awards and prizes received, but this indicator is again not linked to the distribution of funds.

Election to learned academies and academic professional associations

This non-bibliometric indicator is taken to reflect the esteem or standing of individuals by their academic peers in the form of election to learned academies or academic professional associations. To date this indicator has not been directly linked to the distribution of research funds either in Australia or internationally.

Office bearers in learned academies and academic professional associations

This non-bibliometric indicator is thought to be a sign of the esteem or standing of individuals in the eyes of their academic peers in the form being elected as office bearers in learned academies or academic professional associations. Again, this indicator has not yet been linked to the distribution of research funds either in Australia or internationally.

Conferences

Invitations to participate in conferences are seen to represent either the quality of research or the standing of an individual among their academic peers.

Contributions to conferences have been used as an indicator linked to the distribution of research funds in Australia in the form of refereed conference papers in the DEST publication category E1 'peer reviewed papers presented at a conference of national or international significance and published in its proceedings'. These contributions are not weighted in any way. The ARC collects data in its end of grant reports which includes invitations to present conference papers and lectures (particularly keynote addresses), but it is not used in the grant determination process. Since 2001, the NHMRC has started employing new measures to gauge 'Increased capacity of Australian researchers' including keynote addresses, presentations and speaking engagements. This data is one of a number of measures used to rate the record of research achievement of grant applicants in one of its schemes. New Zealand does link conference contributions to national research funding in the form of keynote presentations defined as 'any plenary/invited written or oral presentation delivered at a recognised science forum.'

Service to journals

This indicator is thought to demonstrate the esteem or standing of individuals among their academic peers through being invited to serve on journals in a variety of roles, ranging from editing and editorial board membership to refereeing papers and reviewing books.

Service to journals has not been linked to the distribution of research funds either in Australia or internationally. The Netherlands includes this indicator in its national assessment of research, but this evaluation is not tied to the funding process. In 2005 an attempt at ranking (rather than banding) journals by the UK Arts and Humanities Research Council was abandoned in the face of resistance from the academic community who objected on a variety of grounds to the AHRC generating lists of the top ten 'most significant and important' journals in ten subject areas.

Visiting fellowships

The number of overseas visiting fellows a department receives is taken as a marker of a department's international standing. Conversely, the number of invitations extended to a department's members to be visiting fellows in overseas universities is taken to show the regard in which these people are held internationally.

Data on incoming and outgoing visiting fellowships has not been linked to the distribution of research funds in Australia. The ARC collects data in its end of grant reports on visiting professorships/fellowships/other appointments overseas. In its evaluation of university research performance, Finland collects data on the number of foreign visitors per year a department receives.

3.2 Impact indicators

Quantitative impact indicators represent a novel approach to assessing the effects of research beyond the academic realm. While these have rarely been investigated in the research funding context, the RQF places impact indicators firmly upon the agenda for an Australian research assessment exercise, and in OECD countries there is increasing desire in policy circles to quantify research impact. It is believed that impact indicators will benefit research in applied and policy-relevant areas, and in this respect are particularly suited to the social sciences. For this pilot study all data has been extracted from the annual reports of the participating units.

Within Australia the ‘impact’ of research has largely been gauged in terms of commercialisation and technology transfer, and the Publicly Funded Research Agencies (PFRAs) AIMS, ANSTO and CSIRO, together with the ARC and NHMRC, routinely collect this information. The impact indicators selected for this pilot study shift the focus towards quantitative measures of policy-relevance, which we hypothesise are suited to applied social science generally, and economics particularly.

Government commissioned reports and consultancies

Research undertaken by units which is contracted by government bodies is taken to reflect the direct relevance of a unit’s work. This has not been operationalised as a specific indicator in Australia or internationally, although it can feed into the more generic measure of research ‘linkages’.

Research linkages with government bodies

Research linkages with government bodies are seen to indicate the relevance of applied research to political practitioners. Internationally, the Netherlands’ national assessment of research records ‘user’ linkages but this indicator is not tied to the distribution of funds, while New Zealand does connect such engagement to the allocation of funds albeit in a rather primitive manner.

Advice to Australian government bodies

The provision of expert advice to government bodies is a marker of the esteem in which the opinions of an academic or group of researchers is held, and the importance attached to their work. The Australian PRFAs all gather information on advice given to government. For example, CSIRO collects data on written submissions to government, verbal reports and presentations, advice given, memberships of government committees and invitations to briefings. ANSTO seeks information on contributions to international policy developments. These agencies are increasingly beginning to consider community benefits of research, which marks a further direction impact indicators could take.

4 THE PROCESS

The pilot study was conducted within a seven week period. The first step was to identify economics groups or departments which represented theoretical, applied and policy-driven research, and with a research only or a teaching focus. Five ‘units’ were selected which covered these profiles, one of which was multidisciplinary and highly policy-oriented. The ‘units’ also represent a mixture of junior and senior teaching and research staff, whose average number is taken from annual reports for 2001, 2002, 2003 and 2004.

	Av. staff	Unit Description
Unit 1	15	Full-time research (applied and policy-oriented)
Unit 2	27	Traditional teaching focused
Unit 3	41	Traditional teaching focused
Unit 4	21	Full-time research (basic, applied, policy-oriented)
Unit 5	13	Full-time research (basic and applied)

In week one we met with our Reference Group, comprised of six leading economists drawn from the units being studied. The aim of this initial meeting was to encourage some brainstorming about the types of indicators we could test. We explained the purpose and context of this pilot study, and reinforced the point that REPP did not necessarily support these indicators: we were sure that many would be rejected outright but nonetheless wished to rigorously test them because they were being proposed as measures which could feed into research evaluations and hence affect the distribution of research funds. On the other hand, there were some indicators which we believed held promise and could potentially be developed further.

We stated that the aim of the pilot study was to look at novel measures and what data it was feasible to collect and assess. To this end we were not trying to be fully comprehensive, so wished to avoid standard bibliometric measures that had already been ‘done to death’. We also emphasised the point that we were not interested in producing rankings, and the focus would be on the performance of the indicators rather than the performance of units. REPP guaranteed anonymity, and undertook that no unit or individual would be publicly identified in the course of this research.

Week 1	Five economics ‘units’ identified Reference Group meets Potential indicators narrowed down
Weeks 1 – 7	Data collection from REPP databases, and annual reports Data analysis
Week 7	Reference Group meets Testing the indicators: ‘attacking the measures’

We presented the group with a range of quantitative measures we had already identified as possible candidates for study, discussed the type of data we wanted to collect and possible sources, and took suggestions on any additional data sources or indicators we may have overlooked. The intention was always to rapidly collect data remotely and not to inconvenience the units being studied, so the main sources discussed were annual reports, unit web sites and online CVs. As an outcome of this meeting the range of potential indicators to investigate was narrowed down. We undertook to collate all the information and carry out a comparative analysis of the data, and then to present the findings to the group of expert peers who had agreed to reconvene six weeks later with the intention of 'attacking the indicators' to see how these stood up to intense scrutiny. The main point of the pilot was to use discipline-based peer review to examine the comparative data so that the people being assessed can then assess the assessment: we wanted them to 'attack the indicators', something which seemed to appeal very much to economists' sensibilities.

We agreed that each indicator would be examined in terms of:

Design	What period should the data cover? What are appropriate category descriptors? Are there any field specific characteristics that need to be taken into account when designing the indicator? If bands are used, how should these be weighted?
Validity	Does the data measure some aspect of research quality? If 'yes' what aspects of quality does it relate to?
Fairness	Is the indicator equally applicable to: research sub-fields? units with varying age and gender profiles? teaching and research focussed units?
Reliability	Could the choice of time period significantly affect the outcomes? i.e. is the data likely to be volatile between years?
Transparency	If institutions provide the data, can it be independently verified? If 'third party data' is used, is it easy for those being assessed to verify its accuracy?
Use	How should the indicator be used? e.g. in an RAE style assessment or for internal management purposes? Should it be linked to the distribution of research funds?
Data source	Where would the data come from? What level of aggregation would it have to be collected at? How often would it have to be updated?
Cost	If 'third party data' is used, are the costs justified? If institutions provide the data, how heavy are the compliance costs for institutions and/or researchers?
Behavioural impact	What is the likely effect on the behaviour of researchers if the indicator is used? Is this impact beneficial or unwanted?

In weeks one to seven, the REPP Database and the REPP ARC Linkage Project Experimental Database were used to collect data for each the units for the bibliometric indicators. The

EconLit website was also mined. Annual reports, web pages and online CVs were examined as data sources for the non-bibliometric ‘esteem’ indicators and the impact indicators. We found that annual reports gave the most comprehensive data coverage, and chose to use this medium for the study. All units had annual reports available for the years 2000, 2001, 2002 and 2003. Comparative tables were then produced to present to the Reference Group for their scrutiny.

Seven economists participated in the ‘attacking the measures’ session, and the data provided stimulated lively discussion. The peer assessment of these indicators is reported in the next section of this report. We were concerned that the participants may have become defensive about the performance of their particular units, but this worry was unfounded as the discussion focused upon generic issues and the implications of tying the indicators reviewed to the distribution of research funds.

5 RESULTS

5.1 Quality Indicators

5.1.1 Standard bibliometric indicators

To help in the assessment of the novel indicators we presented, contextual data was provided in the form of three standard bibliometric indicators presented below: number of DEST publications, citation per publication rates, and journal publications weighted by journal impact factor. The analysis was limited to the five economics 'units' used in the pilot study, and covers three years: 1995, 1997 and 1999. Those three particular years were chosen to match the period covered for citations to non-source publications (see 5.1.2 below).

There are two sources of data. The first is constructed from data files purchased direct from the Institute for Scientific Information (ISI) – the REPP ISI Database. It contains all Australian publications in ISI's three main indexes – the Science Citation Index, the Social Science Citation Index, and the Arts and Humanities Citation Index. It is updated annually, and currently covers the period from 1981-2003. The addresses on publications are 'cleaned' by REPP, unifying all variants of an institution, faculty or department under one standardised address.

The second data source is REPP's ARC Linkage Project Experimental (ALPE) Database, which was constructed from the bibliographic details supplied by 26 universities for their DEST publications. Many universities also provided full details of additional publications collected for internal purposes. The experimental database contains data for an eight year period, 1995-2002, though not all universities were able to supply data for the full period. All publications are linked to departments and/or schools.

Number of DEST publications

This indicator comprises a simple count of publications – in this instance, limited to the four DEST categories: a scholarly book produced by a commercial publisher (A1), a chapter in a scholarly book produced by an international publisher (B1), an article in a scholarly refereed journal (C1) and a peer reviewed paper presented at a conference of national or international significance and published in its proceedings (E1). The data was extracted from REPP's ALPE Database.

Table 1: Number of DEST Publications

	A1	B1	C1	E1	Total
Unit 3	8	62	173	29	272
Unit 4	6	57	92		155
Unit 5	2	14	46	4	66
Unit 2	1	9	41	1	52
Unit 1	0	5	17	5	27

This indicator did not appear to favour either traditional teaching units (2 and 3), or full time research units, all of which engaged in some research that could be classified as 'basic'. However, Unit 1, primarily focused on applied and policy-oriented research, had relatively few publications that satisfied the restrictive definitions that applied to these four categories.

Citation per publication rates (ISI journals only)

This indicator is a simple citation per publication rate. The REPP ISI Database was interrogated to identify the number of citations that each of the ISI journal publications from the economics 'units' attracted between the time of publication (1995, 1997 and 1999) and 2002, and the average rate per publication was calculated.

Table 2: Citation per publication rates (ISI journals only)

	No. Publications	No. Citations	Citations per Publication
Unit 5	20	111	5.55
Unit 3	35	153	4.37
Unit 4	35	113	3.23
Unit 1	1	3	3.00
Unit 2	24	51	2.13

The introduction of citation counts, and calculation of average citation rates, puts a different complexion on the performance of units. Unit 5 has a citation rate well above that of other units. On face value, the measure does not appear to favour teaching units over full-time research units. However, in all cases the counts are based on a very small number of publications, and a small proportion of the units publication output. Unit 1 is virtually 'invisible' in this indicator.

Journal publications weighted by journal impact factor (ISI journals only)

This indicator uses publication counts weighted by journal impact factor (i.e. a journal's *potential* or *expected* citation impact). The impact factor data was obtained from ISI's *Web of Knowledge*, and applied to publication data extracted from the REPP ALPE Database.

Table 3: Journal publications weighted by journal impact factor (ISI journals only)

	No. Publications	Tot IF	Av IF	% in ISI	Total Pubs
Unit 5	25	25.19	1.01	56	45
Unit 2	21	15.17	0.72	51	41
Unit 4	35	20.53	0.59	38	92
Unit 3	25	14.82	0.59	14	173
Unit 1	0			0	17

Using the impact factor of the journals in which publications appear produces a quite different ranking to that created by counting the *actual* citations they attract. It highlights the dangers of relying on journal data rather than publication data to assess units, for it ascribes to all articles in a journal an average score. Thus it under-values articles that attract more citations than the journal average, and over-values those articles that attract few citations.[6]

[6] A comparison of publication numbers between Tables 2 and 3 highlights the problem of using different data sources to construct different indicators. Where large numbers of publications are involved this is not always an issue but, as in the data shown in the two tables, the discrepancies can be quite large at lower levels of aggregation – only 20 ISI publications for Unit 5 were identified in REPP's ISI Database, but 25 were identified

Response to standard bibliometric indicators

The three standard bibliometric indicators were provided as benchmarks against which to compare the non-standard indicators, and indeed the 'esteem' indicators and impact indicators. While we did not have an in-depth discussion of this data, our workshop participants did not express any major surprise at the findings. Nor did we seek to assess them separately, as we did the other measures. However, they did give rise to a general discussion of the efficacy of bibliometric indicators, and the comments elicited are worth presenting at this juncture.

While citation indicators were not rejected, there was some concern that the length of time it took for articles to be published meant that there needed to be a fairly long citation window. The example was given that a paper citing another article published in 1999 might not appear in print for another two years. Even if the publication has an 'immediate' impact, it would not be seen in the citation data for some time. A minimum 5 year window was recommended.

It was felt that ISI coverage was very patchy in economics – one economists believed that only 130 of 740 economics journals were in the indexes.

The importance of working papers in the field was likely to impact on citation counts. Colleagues might find and cite working papers before they appeared in their final journal form, and these citations would not be counted as they could not be ascribed to the relevant journal article.

There was a consensus that publication counts should be calculated on a per capita basis, and important contextual information to put alongside bibliometric data was the proportion of staff who did not publish at all. There was even a desire to see citation per person calculations alongside citation per publication calculations because there was a belief that the latter did not fully account for size adjustment. One methodological problem raised, but not resolved, was whether publication (and citation) counts should be fractionated by the number of authors.

Because of the extremely skewed nature of citation data, the calculation of both medians and means was advocated.

There was a concern about how multidisciplinary work would be affected by discipline-based citation measures. The value of such work should not be discounted or under-valued because it did not fit neatly within disciplinary boundaries.

Some participants believed that citation data could be manipulated by extensive self-citing, though others believed editors could reduce the occurrence of this with careful scrutiny of reference lists in the review process.

One idea proposed was to weight citations by the prestige of the journals citing the paper. This raised concerns about the potential to skew counts against inter-disciplinary publications, and the cost of incorporating this additional data into the measure. While this is routinely done at the University of Leiden, few other bibliometricians have access to the data required to do this.

---

in its ALPE Database. Discrepancies were primarily due to inaccurate bibliographic details and the lack of an Australian address – both issues that would prevent an article being identified in REPP's ISI Database.

5.1.2 Non-standard bibliometric indicators

Data for two non-standard bibliometric indicators were presented for peer scrutiny: (1) citations to all publications, and (2) publication counts weighted by journal or publisher prestige.

Citations to all publications (i.e. not just ISI journal articles)

Data cover 1995, 1997 and 1999 publications, and citations were counted to the end of 2002. This data has been collected as part of REPP's Linkage project, where we have combined with the Centre for Science and Technology Studies (CWTS) at Leiden University to mine their full ISI database for citations to non-source publications, thus extending the coverage of publications for which citation data can be assembled. The analysis was restricted to the four DEST publication categories. Relevant data for the units in our study was extracted from REPP's ALPE Database and citations to 'non-source' publications were obtained from the CWTS database. Citations per publication were then calculated on this expanded set of publications, rather than being restricted to articles appearing in journals indexed by ISI.

Table 4: Citation per publication rates including non-source items (all DEST publications)

	No. Publications	No. Citations	Citations per Publication
Unit 5	53	112	2.11
Unit 4	156	196	1.26
Unit 2	53	61	1.15
Unit 3	269	296	1.10
Unit 1	28	11	0.39

One effect of the extension of citation analyses to include non-source items is the considerable increase in the number of publications covered in the analysis. The bulk of additional coverage related to journals not indexed by ISI, but also included book chapters, and a small number of books. The fourth DEST category – conference publications – was not visible in the database for these units (or indeed for any other economics unit). However this increased coverage was not always matched by a marked increase in the citation counts. Units 1, 2 and 5 increased only minimally, while the counts for units 3 and 4 almost doubled. As a result, the ranking of units changed significantly, though unit 5 remained at the top, showing that this indicator benefits research-only units, followed by teaching units and the policy-focused unit.

Not surprisingly for a group of economists, discussion focussed on a cost-benefit analysis of the measure. One participant was not convinced that, for economics, the amount of cost/effort required to construct the measures would be worth the return. In general, participants questioned the introduction of a measure which included some relatively unimportant publication types for their discipline (book chapters and conference publications) but could not cover one important mode – working papers.

Though noting comments on what types of publications are the most important for economists, incorporating non-source items in any citation analysis was predicted to raise their 'value', and this was seen as a positive effect.

Publication counts weighted by prestige (publisher or journal)

Two examples of weightings, both based on the use of five prestige bands, were used, but many alternative systems could be constructed. Three DEST categories are covered – books, book chapters and journal articles. Data covers 1995, 1997 and 1999 publications.

For books, we followed the CERES[7] classification of publishers, and for journals we used rankings compiled by Dr Anne-Wil Harzing of the University of Melbourne,[8] supplemented by ISI impact data for those ISI journals not contained in the Harzing lists. Copies of these lists were provided for the 'attacking the measures' session. Publishers not listed in the CERES scheme could fairly confidently be allocated to a band using the compiler's specifications of how the bands were constructed, however we did not have the knowledge or expertise to allocate missing journals to bands.

In the following tables we have classified publications into bands A to E, following the schema of a number of rankings contained in the Harzing lists. Typically an A ranked journal (i.e. in the highest band) is described as 'top journals with world-wide distribution and readership' and generally covered the 'entire scope of a discipline' and had a strong refereeing process. Journals at the lower end of the scale were typically 'lower level national journals'.

Table 5: No. Publications

Rank:	A	B	C	D	E	Tot
Unit 3	9	57	67	47	27	207
Unit 4	16	54	23	30	10	133
Unit 5	16	21	18	4	1	60
Unit 2	4	22	9	6	0	41
Unit 1	0	1	3	13	8	25

Note: books are given a weight of 5

Table 6: Publication prestige weights (1)

	Weighted publication score	Average publication score
Unit 5	227	3.8
Unit 2	147	3.6
Unit 4	435	3.3
Unit 3	595	2.9
Unit 1	47	1.9

Note: weights applied are A=5, B=4, C=3, D=2, E=1

[7] Centre for Resource Studies for Development, University of Utrecht: http://ceres.fss.uu.nl
[8] http://www.wu-wien.ac.at/usr/h98d/h9851615/journalquality_harzing.pdf

Table 7: Publication prestige weights (2)

	Weighted publication score	Average publication score
Unit 5	2288	38.1
Unit 2	1093	26.7
Unit 4	3319	25.0
Unit 3	2858	13.8
Unit 1	52	2.1

Note: weights applied are A=100, B=30, C=3, D=1, E=0

This measure produced rankings the same as those obtained when ranking ISI journal publications by impact factor (see Table 3). It is worth noting that both are based on the assessment of journals rather than publications. The effect of applying different rating scales can be clearly seen by comparing Tables 6 and 7. While the ranking of institutions remains unchanged, there is a clear difference in the relative performance between institutions. With a higher weighting for the top two bands, Unit 5's average score moved clear of those for Units 2 and 4, and Unit 3 dropped well below the top three institutions.

Participants rejected the weight of 5 given to books. They believed that, at least for economics, it should have the same weight as any other publication.

Participants were keen to see how closely the proposed rankings aligned to journal impact factors. If these are closely correlated, do we need two sets of indicators? Alternatively, if the correlation was not so strong, then it would be useful to demonstrate that the most prestigious journals are not necessarily the ones with the highest impact factor.

Weightings similar to those in Table 7 were thought to be less open to manipulation than those in Table 6. They gave a much higher weighting to those journals or publishers judged the most prestigious. One participant compared the different weightings by noting that in the first version (used in Table 6), one article in *American Economic Review* was worth two in *Economic Record*; while in the second version (used in Table 7), one publication in *American Economic Review* was worth 30 in *Economic Record*.

Some time was taken in discussing how to determine the weighting scales. In the previous example, it was felt the first option was 'soft', with more differential needed between the bands. However, it was felt the second option went too far in the opposite direction. For journals, an alternative weighting scheme was suggested, using the average impact factor for each group. This would be easy to calculate and has some basis for application.

It was suggested that if there are fewer bands, e.g. three rather than five, there will be less disagreement around the margins.

There was general support for this measure, particularly for its ability to include non-ISI and more local and applied journals, though the difficulty of classifying these additional journals was acknowledged.

There were also residual concerns about problems of multidisciplinary and interdisciplinary research. The multidisciplinary nature of Unit 1 was an example of this problem, as this unit contained several anthropologists. It was essential that in any assessment, quantitative measures would need to be sensitive to this issue, perhaps by widening the discipline base of the assessment process.

In a similar vein, participants raised question about an underlying premise of any ranking scheme for economics journals – what are we trying to measure? In particular, for multidisciplinary journals, is it impact on all academia, or only on economists?

Working paper views and downloads (EconLit)

The data presented covered the 12 month period June 2004 to May 2005, taken from LogEc which collects online access statistics from several services which use the RePEc data set, the largest online collection of economics working papers, journal articles and software components, incorporating: EconPapers, IDEAS, New Economic Papers, NetEc and Socionet. The statistics are updated monthly. To avoid double counting views and downloads, LogEc keeps track of the originating IP-number of each access and counts only one access to a specific resource. Data is restricted to registered authors only, which is activated on an individual, voluntary basis through the RePEc author service, so we report on a self-selected sample.

We have chosen not to present the underlying data we assembled on working paper views and downloads as individuals and units can be identified too easily, and this would run contrary to our undertaking to preserve anonymity. Instead we have produced a summary table showing the distribution of views and downloads between the units. However, figures for individuals were provided as contextual information for the 'attacking the measures' session.

Our focus of attention was drawn to the importance of working papers amongst economists, although the discussion also covered the idea of using downloads more generally as potential indicators of research quality. We were struck by the fact that for working papers the units collectively attracted over 187,000 abstract views and 22,908 full downloads; and for journal papers 36,537 abstract views and 6,181 downloads.

Table 8: Percentage of Working Paper Views and Downloads by Unit

Unit	Registered Academics	Working Papers Abstract views %	File Downloads %
4	4	35	36
5	3	31	36
3	16	14	15
2	6	16	10
1	3	3	3

When distributed between units, the two full-time economics research units attract the most working paper views and downloads, followed by the traditional teaching units and the full-time multidisciplinary applied and policy-oriented research unit. However, coverage is clearly not complete as there are many academics are not registered. We noted that there were several 'star' performers attracting around 8,000 abstract views, and individuals who recorded above 16,000 or 36,000 or 40,000 similar views. There were also a couple of economists who had over 1,000 working paper downloads, and individuals with over 2,000 or 3,000 or 4,000 downloads.

The peer discussion reinforced the centrality of working papers to communicating economics research. It was felt that this was because research had its most impact in its initial stages, so with a two year lag to submitting an article and journal publication, working papers act as a swifter route to transmit and exchange ideas. Indeed, economists routinely submit a journal paper and simultaneously release this as a working paper (so one would need to take care to not double-count these publications).

It was felt that to be rewarded for the number of downloads of working papers is a good incentive. But there were some concerns that download numbers could somehow be manipulated, although this was counterbalanced with the view that the numbers of downloads are so large that this is unlikely to prove a significant problem.

The view was expressed that downloads should not be tied to funding as they are too imperfect: they reflect bias towards popular areas and famous names. The consensus was that downloads are perhaps best understood as an indicator of 'interest', or even as a measure of research activity rather than research quality.

Summary

While the three novel bibliometric indicators did not present any surprising results, there was some support for at least one of the measures. The main benefit seen was an increased coverage of the total output of the discipline, and due weight being given to important journals not indexed by ISI. Discussions highlighted the view that journal publications and working papers were the only important outlets for academic research, with the possible addition of books, restricted to those published by the major academic presses or a few prestigious commercial publishers.

The extension of bibliometric analysis to incorporate citations to non-source items was not seen as a particularly useful development. While it did increase the coverage of output beyond ISI indexed journals, it did not include one of the most important modes of communicating research in this discipline – working papers. Insufficient benefit was gained to warrant the cost of such an exercise.

Weighting publication counts by prestige (either of the publisher or of the journal) was seen as a useful extension of a standard bibliometric measure (journal counts weighted by ISI impact factor). It increased the coverage of output in the discipline by incorporating books, book chapters, and articles in non-ISI journals. While the measure still needed further development in terms of the number of bands, the weightings to be given to each band, and methods for assessing non-ISI journals, participants gave general support for this measure.

The centrality of working papers to economics is generally not acknowledged in bibliometric studies. We were struck by the sheer number of views and downloads from the EconLit website. However, participants questioned the validity of the data for the stated purpose: they were reluctant to see it inform funding decisions as they felt the data could be heavily influenced by factors unrelated to quality.

5.1.3 Non-bibliometric indicators

Several non-bibliometric or 'esteem' indicators were explored and data is presented below for honours, awards and prizes; election to and roles within learned academies and academic professional associations; conference activities; service to journals; and visiting fellowships. All data was drawn from the annual reports of participating units for the years 2000, 2001, 2002 and 2003.

Honours, awards and prizes

Data is presented for the four combined years 2000 to 2003, and aggregated to the level of research unit. 'Honours' are divided into 'general' (e.g. various medals bestowed by government or professional academic associations); and honorary professorships and honorary degrees awarded by Australian and overseas universities. ‘Awards and Prizes’ are split into those received for research (e.g. research excellence recognised by an overseas university, an overseas state government, an Australian or an overseas professional academic association) and publication (e.g. for papers or books, awarded by academic professional associations in Australia and overseas, international academic journals and overseas universities. These may or may not have been awarded on the basis of open competition). Lists of all honours, awards and prizes were supplied for the ‘attacking the measures’ session.

Table 9: Number of Honours, Awards & Prizes Received

	Honours			Awards & Prizes	
	General	Professorship	Degree	Research	Publication
Unit 5	10	0	2	0	2
Unit 2	1	0	0	1	2
Unit 3	0	2	0	0	1
Unit 4	1	0	0	0	1
Unit 1	0	0	0	0	0

We found that the distribution of honours, awards and prizes was sparse over the four year period. ‘General’ honours mostly took the form of various Centenary Medals and Order of Australia medals. There were two honorary professorships at overseas universities, and two honorary degrees from overseas universities. Awards for research and publication mostly came from academic professional associations. Unit 5, a full-time research unit clearly received the bulk of honours, awards and prizes.

In the ‘attacking the measures’ session, the consensus was that these ‘esteem’ measures were precisely that: measures of esteem and not of research quality. There were many caveats, for example, government medals were not awarded on the basis of cited research; as one participant put it, the title of honorary professor often ‘just kind of happens’ and tends to denote nothing more than an adjunct position; and distinguished author status, listed as a publication-based award, is given automatically for publishing a certain number of papers in the same journal within a set amount of years.

It was generally felt that awards and prizes for research were acceptable performance measures.

There was great discomfort with the idea of ‘esteem’ being distilled into indicators, and with funding being allocated on the basis of honours, awards and prizes, although it was felt that this kind of information could legitimately be fed into some kind of departmental contextual statement to inform a peer review process.

Election to learned academies and academic professional associations

The data covers 4 combined years: 2000, 2001, 2002 and 2003, and is aggregated to the level of research unit. There is no double-counting. Election to learned societies and professional associations was recoded when initials of the institution was given after people’s names in annual reports: however these elections mostly predate the period 2000-2003. ‘Learned Societies’ include being elected fellows of Australian or international academies, institutes and royal societies; while ‘Professional Academic Associations’ include being elected a member of Australian, overseas or international discipline-based bodies and professional institutes.

Table 10: Number of fellows of learned societies and professional associations

	Learned Society	Professional Academic Association
Unit 5	9	5
Unit 4	7	0
Unit 3	4	8
Unit 2	2	2
Unit 1	2	1

The two full-time research units have the most elected fellows of learned societies, followed by the traditional teaching units, and the full-time multidisciplinary policy-oriented unit. A teaching unit has the most elected members of professional academic associations.

Did the peers regard this as an indicator of research quality? As we shall see, they again took this to be a marker of esteem rather than research quality.

Conversation focused largely on elections to learned societies, and while data on learned academies was seen as an acceptable esteem measure, data on professional associations was not. There is a problem with this data, which is that people could have been elected to academies or associations at any time in the past (22 of the 24 elections to learned societies were prior to the year 2000), so this indicator reveals little about recent achievements, or as one ‘attacking the measures’ participant put it, reflects a stock of esteem and not the flow of esteem. The group of peers felt that people with the title FRS were probably ‘old men’. These two factors combined raise concerns that this indicator is biased against younger economics departments and young researchers, which would raise problems if esteem indicators were used to produce hierarchies and linked to the distribution of funds.

Another point noted was that because the count was of memberships, a unit’s figures could all be for one person. Also, there are many stories of perceived injustices about people elected and not elected to academies and associations: elections may be awarded or declined on the basis of many factors not necessarily related to research quality so there would be little likelihood of persuading the academic constituency to accept this indicator.

Office bearers in learned academies and academic professional associations

Data was collected for service to learned societies and academic professional associations for the combined years 2000, 2001, 2002 and 2003, and aggregated to the level of research unit. There was no double-counting.

Table 11: Number of office holders in learned societies and professional associations

	Learned Society	Professional Academic Association		
	Treasurer	President	Vice-President	Treasurer
Unit 5	1	3	3	0
Unit 3	0	4	1	0
Unit 2	0	2	0	0
Unit 1	0	0	0	1
Unit 4	0	0	0	0

We found that this indicator produced small numbers, although there were several instances of being elected president or vice-president of a professional academic association. One full-time research unit had three of each, while the other had none. A traditional teaching unit with a heavy involvement in professional academic associations recorded three presidencies and one vice-presidency, and the other traditional teaching unit had two presidents.

However, little currency was attached to these positions by the peers. For example, it was felt that someone may be elected a treasurer of a learned society because they are good at doing the books and not for academic reasons. One participant mischievously suggested that being an office bearer in a professional academic association was a negative indicator of research quality!

Conferences

Data was collected for the combined years 2000, 2001, 2002 and 2003, and aggregated to the level of research unit. This was a very time-consuming task, so data was only collected for Units 3 and 5, one teaching-focused and one research-focused unit. Contributions to conferences included major national and international economics annual meetings, smaller subfield-related events, conferences in other academic fields, and practitioner-related events. These are divided into international and national conferences.

Table 12: Number of contributions to national and international conferences

	International Conferences									National Conferences								
Unit	Invited Keynote	Invited Paper	Keynote	Convenor	Chair	Paper	Discussant	Panel Member	Poster	Invited Keynote	Invited Paper	Keynote	Convenor	Chair	Paper	Discussant	Panel Member	Poster
3	0	1	2	0	4	22	2	0	0	0	1	0	2	0	38	2	1	1
5	4	2	5	0	0	3	0	0	0	2	9	5	1	0	0	0	0	0

Note: papers and keynote talks are marked as 'invited' where this is mentioned in annual reports, and these reports did not state whether papers were refereed.

The information provided in annual reports did not allow us to consistently discern whether papers were invited or not, or if they were refereed. We found that on the basis of the information given, Unit 3 – the larger teaching-oriented unit – provided the higher number of papers in Australia and internationally, while Unit 5 (research only) presented more invited papers and gave more keynote talks.

It was felt that this indicator had a trivial relationship to research quality: in economics, conferences are not regarded as a reflection of research quality as very few papers are refereed. The real test was felt to be whether the paper is turned into a publication – and if it did then this would lead to double-counting the same piece of work. However, being a keynote speaker was seen as an 'esteem' measure worth pursuing, although as part of a context statement rather than an indicator in its own right.

It was believed that this measure could be easily manipulated, as people could set up their own conferences and engineer 'invitations' to present papers. Indeed, one participant confessed to listing all their conference papers as 'invited' following the logic that at some stage they are in fact asked to give the paper.

It was also noted that this is an input rather than an output indicator.

Service to journals

Data was collected for 2000, 2001, 2002 and 2003 separately, and aggregated to the level of research unit. The journals covered are largely peer reviewed academic international or Australian publications, and are mostly in the discipline of economics, but many interdisciplinary journals and serials which belong to other fields are represented along with journals in numerous sub-fields and publications aimed at practitioners.

Table 13: Number of unit members engaged in service to journals 2000

	Editor	Co-Editor	Guest Editor	Editorial Board	Reviews Editor	Referee	Book Reviewer
Unit 4	1	1	0	23	0	25	0
Unit 2	0	5	0	8	0	36	0
Unit 3	0	5	0	19	0	12	0
Unit 5	1	5	0	10	0	0	0
Unit 1	0	1	0	2	0	8	0

Table 14: Number of unit members engaged in service to journals 2001

	Editor	Co-Editor	Guest Editor	Editorial Board	Reviews Editor	Referee	Book Reviewer
Unit 4	2	3	0	25	0	46	0
Unit 5	1	5	0	12	0	29	0
Unit 3	1	5	0	18	0	0	0
Unit 2	2	3	0	8	0	0	0
Unit 1	0	1	1	1	0	8	0

Table 15: Number of unit members engaged in service to journals 2002

	Editor	Co-Editor	Guest Editor	Editorial Board	Reviews Editor	Referee	Book Reviewer
Unit 5	1	2	0	18	0	27	0
Unit 4	3	6	0	12	0	19	1
Unit 2	0	2	0	0	0	33	0
Unit 3	4	9	0	20	0	0	0
Unit 1	0	1	0	2	0	12	0

Table 16: Number of unit members engaged in service to journals 2003

	Editor	Co-Editor	Guest Editor	Editorial Board	Reviews Editor	Referee	Book Reviewer
Unit 5	1	2	0	20	0	53	0
Unit 4	3	6	0	12	1	23	1
Unit 3	1	6	0	26	0	0	0
Unit 2	2	1	0	1	4	21	0
Unit 1	0	1	0	2	0	10	0

We find that over the four year period the two research-focused economics units (4 and 5) provide most service to journals, along with the teaching-oriented units (2 and 3). This involvement reflects editing journals, being an editorial board member, and reviewing papers. However, the numbers are volatile between years reflecting the changing patterns of reporting within annual reports.

Opinions on the potential value of this indicator were mixed. For example, one view was that refereeing is part of normal academic workloads, and so should be excluded from the measure, and another view was if an assessment of quality takes into account workloads, then this measure might be an important indicator. However, grappling with who are part-time versus full-time researchers will present assessors with problems.

It could be argued that all these roles are esteem rather than quality measures, although editing (viewed as a low indicator of quality) was nonetheless 'worth' more than writing a book review. One participant drew from their own experience to explain that being a member of an editorial board is an esteem measure and not a quality measure, because they did not actually do anything in their own role as editorial board member.

One conclusion was that service to journals was a positive but low measure of research quality, and simple counts should be weighted to type of service and the quality of the journal to which service is provided. Another conclusion was that this indicator muddles quality versus workload, and that if we followed this line we might as well include 'teaching' as an indicator of research quality, which would obviously be absurd.

Visiting Fellowships

Data covers the combined years of 2000, 2001, 2002 and 2003, and is aggregated to the level of research unit. Only international ingoing and outgoing fellowships and other appointments were recorded, and their duration varied from one month to one year. Double-counting was excluded.

Table 17: Incoming international visiting fellows and other appointments

	Visiting Professor	Visiting Fellow	Visiting Senior Lecturer
Unit 5	0	85	0
Unit 4	0	64	0
Unit 3	8	18	1
Unit 2	0	14	0
Unit 1	0	9	0

Table 18: Outgoing international visiting fellowships and other appointments

	Visiting Professor	Visiting Fellow
Unit 2	0	3
Unit 4	1	1
Unit 5	0	2
Unit 1	0	1
Unit 3	0	1

Incoming international visiting fellows are clearly distributed most amongst the traditional research-only units, while outgoing international visits present very small numbers.

This measure was not taken very seriously, and it was noted that the number of visitors units receive depends upon the size of budgets set aside for this purpose.

Summary

When examining the results produced by the non-bibliometric 'esteem' indicators, we find that these benefit most the two economics research-focused units, followed by the two traditional teaching units, with the multidisciplinary and policy-oriented unit least well served. There are no surprises in this, and while on the one hand this suggests that these indicators correlate with research activity, they also mirror this activity and would reward the most research focused units for being the most research focused units. It was felt that these indicators reflect esteem rather than actual research quality, and sometimes reflect research-oriented workload.

In this light we may ask if these indicators tell us anything new beyond bibliometric approaches, or if they may be used as a replacement for or an addition to standard and non-standard bibliometric measures. The 'esteem' indicators are less volatile than the bibliometric measures presented, although they similarly favour the research-led units.

However, peer review found these measures to be less credible than the publication-based indicators, and to reflect of esteem and not quality. It was also felt that these 'esteem' indicators duplicate what can be done with bibliometric measures, and are backward looking, and hence presenting problems for early career researchers and young departments. Their use as stand-alone 'metrics' linked to the distribution of research funds was roundly rejected, although this information was seen to be potentially useful in informing context statements for peer review panels in a research assessment exercise.

5.2 Impact indicators

The idea of developing impact indicators to inform decisions about the allocation of resources for publicly-funded research is relatively new, and while increasingly attractive to policymakers is not well thought out. There are several ways in which research impact can be imagined, and in the context of the Australian RQF process has been defined as the 'broader social, cultural, economic and environmental impact or usefulness' of academic research. The impact indicators presented below are a tentative step into a territory where there is much work yet to be done, and these focus on policy advice, consultancies and linkages with 'users' of economics research beyond academe, and as has been said in Section 3.2 represent the sort of indicator that is being explored by funding agencies in Australia. All data was drawn from the annual reports of participating units for the years 2000, 2001, 2002 and 2003, and there is some double-counting.

It is important to note that there are differences in what information is and is not given in annual reports, so where information is not supplied it cannot be counted. This was a 'quick hit' exercise intended to supply some data to prompt discussion on the potential value, or otherwise, of impact indicators, problems which may be encountered in operationalising such measures, and whether such impact indicators should be linked to research funding. In this light, a more thorough exercise can be conducted, and while the results presented below may not be fully accurate they represent a general picture of what impact data may look like.

Government commissioned reports and consultancies

The number of commissioned reports and consultancies per year were counted (rather than presenting the cash values of such consultancy work).

Table 19: Commissioned reports and consultancies

	Commissioned Reports/Consultancies
Unit 1	30
Unit 4	5
Unit 3	3
Unit 5	0
Unit 2	0

Bearing in mind the caveats about reporting consultancies, we find a small number of these in one research-oriented unit and one traditional teaching unit. However, Unit 1, which focuses on policy-oriented and multidisciplinary research, reports a significant number of commissioned reports and consultancies.

The peer assessment of this indicator revealed that using the number of specific consultancies as a measure overlooks units with core funding from government, such as Unit 1 and a subset of Unit 5, and it was argued that core funding should be included and given a greater weighting. It was argued that weightings should also be given for funds won through competitive processes and for recognised quality in a policy area.

It emerged that Unit 3 does not routinely provide this type of information in its annual report.

Research linkages with government bodies

A count was made of the number of reported research linkages or research collaborations with government in Australia.

Table 20: Research linkages with government

	Research linkages with Govt
Unit 1	26
Unit 3	8
Unit 5	7
Unit 2	3
Unit 4	2

Several instances of research linkages with government were reported and were more evenly distributed between the economics units focused on research and teaching. Again, it is the policy-focused, multidisciplinary unit which stands out.

Data was collected on Australian linkages only, and Units 1 and 4 would have benefited if data – which was available – was collected on international linkages.

Advice to Australian government bodies

Counts were made of verbal and written submissions to policy processes and government enquiries, and of invitations to give advice to various government bodies within Australia.

Table 21: Advice to Australian government

	Advice to Australian Govt
Unit 1	73
Unit 5	43
Unit 4	34
Unit 2	16
Unit 3	0

We find that the teaching-focused units are less involved in advising, with research-focused economics units in the middle ground, and the policy-oriented and multi-disciplinary unit provides more instances of advising government.

In the 'attacking the measures' session we discovered that Unit 3 does engage in this activity but does not record this information in its annual report, raising again the issue of standardisation of information collected remotely.

A case was made that contributing to Senate enquiries, for example, was very time consuming, and that it was important to consider how much labour goes into such activities, and that this will effect how much journal publication can be achieved.

This was accepted as an indicator of research quality by some, although there was a need to band research linkage with international, national and State governments. One participant humorously described research applied to State governments as 'more like community service', and it was felt that State level data was more an indication of community links, but that at the Federal level it might be a more valid indicator of research impact. There was divided opinion on whether international advice was more closely aligned to research quality.

Summary

Some of the peers felt that 'impact' indicators were in fact measures of research quality, while for others these were indeed valid measures of research impact. The importance of the uniformity of reporting was emphasised.

An important question is whether impact indicators tell a story beyond what bibliometric and 'esteem' indicators reveal. The most interesting result was that Unit 1, which was barely visible in the bibliometric and esteem indicators, clearly tops these impact rankings. This inversion was felt to be a trade-off between being an applied and policy-focused unit which spent its time doing consultancy work for various non-academic 'end users' of research versus producing articles for top journals. It was felt that Unit 1 was a special case as it is multidisciplinary, applied and very highly regarded within government, yet suffers in 'quality' and 'esteem' assessment even though it is seen as a very high quality unit with an excellent reputation within and without academe. This and similar units would make ideal case studies for investigating the seeming juxtaposition of quality and impact indicators which we may infer from the data produced for this pilot study, an issue with which any serious assessment of the purpose and value of publicly funded research must grapple.

6 Conclusions

The aim of this pilot research was to obtain and assess some evidence, using a peer-informed comparative process, to identify possibilities for widening the kinds of metrics used to inform research assessment processes, especially for areas of research not adequately represented by standard indicators.

The economists in our pilot study rejected the extension of bibliometric analysis to non-source items as these were unlikely to provide any useful additional information, particularly as one of their two main modes of scholarly communication – working papers – were excluded. With some reservations, they were attracted to the idea of prestige weighted publication counts as it allowed for the inclusion of non-ISI journals, particularly Australian and policy periodicals.

Working paper views and downloads were not accepted as valid measures of research quality, but were seen as a useful indicator activity and hence warranted further exploration.

Esteem indicators were rejected as valid stand-alone non-bibliometric measures of research quality, but were accepted as markers of reputation which can potentially inform context statements about a unit's research.

There was an unresolved debate about whether impact indicators measures do in fact measure research impact or instead gauge some aspect of research quality. There was support for further developing these impact indicators, particularly for assessing policy-oriented and interdisciplinary or multidisciplinary research units.

Overall, we found that there was a tendency for participants to create hierarchies and desire weightings within indicators, which to a large extent rely on publication-based indicators (editor of which journal? was a prize awarded for publishing?), or esteem markers (visited which university?) or other implicit quality-based judgements (was a research linkage with an overseas government, the Federal government or a State government? is funding for research linkages derived from core funding from government/won through a competitive process?)

There was support for three levels of indicators – research quality, esteem and impact, and ensuring the three were separately identified and not confused. Any assessment should therefore make very clear which of the three was being addressed, and separate indicators should be constructed for each. There was agreement that any measures need to kept simple.

7 Next Steps

An important next step is to circulate the first version of this working paper to the pilot study participants and to invite further comments and observations, which will then be incorporated into the next version of this paper, along with other *post hoc* comments we have received.

There are many more novel indicators that are often discussed, and yet to be operationalised, which we wish to investigate. For example, citations in government reports and Hansard, media presence, and Copyright Agency Limited (CAL) data including Public Lending Right and Educational Lending Right fees.

It will be very interesting to extend this pilot work other research areas as diverse as the creative arts, computing and engineering, which are not well served by standard bibliometric indicators.

A further step will be to investigate the use of qualitative measures of research quality and impact, and to test the use of peer and ‘user’ informed qualitative logic modelling as the basis for judgements about research quality and impact (e.g. adapting the HERG/RAND ‘payback methodology’ which has yet to be applied to research in the humanities, arts or social sciences).