The Fate of Computing in Research Performance Evaluations: ERA vs PBRF

Alison Clear
Christchurch Polytechnic Institute of Technology
Christchurch
Alison.clear@cpit.ac.nz

Tony Clear
Auckland University of Technology
Auckland
Tony.clear@aut.ac.nz

Abstract
The prevailing ‘audit culture’ in national governments has seen a global proliferation of research performance evaluation schemes. Most recently the Excellence in Research for Australia (ERA) results have been published. The results from this bibliometrically based ranking exercise provide an interesting comparison with the earlier results from New Zealand’s Performance Based Research Fund (PBRF) exercise. With a focus on the computing disciplines this paper sets these developments in the global context; compares the outcomes under each scheme; the extent to which the prevailing publication cultures have been supported or undermined; the scope for such schemes to render whole sub-disciplines invisible and the potential impacts for the computing disciplines from such exercises.

Keywords
Research Performance Evaluation, bibliometrics, ERA, PBRF

Introduction
The substantial Government support allocated to research in universities leads to a natural concern for value for money on this ‘investment’. Globally government directed funds may account for “over half of all support for university research and are an important share of all public support for R&D” (OECD, 2002, p.21).

A study of Government funding policies (von Tunzelman et al., 2003) identified four main approaches to allocating research funding to universities. The first group of countries used a performance based approach to distribute funds; the second used an indicator other than research evaluation, such as student numbers; in the third group research allocations were ‘open to negotiation’; and in the fourth group research assessment and funding were separated.
New Zealand and Australia have opted for differing variants on the performance based research funding model. In New Zealand funds are allocated based upon the Performance Based Research Fund (PBRF). In Australia a new scheme has been launched to address research funding shortfalls in the existing models, titled the Excellence in Research for Australia (ERA) scheme. This assessment of research links to funding schemes such as the existing Research Infrastructure Block Grant, extended via the Sustainable Research Excellence in Universities initiative (SRE). Bailes (2011, p. 7) notes that access to SRE funding is dependent upon participation in the ERA scheme, and that “some of the funding available under the scheme (1/3 of a total of approx $121 M for 2011) appears to be tied to ERA performance”.

The ERA scheme makes extensive use of bibliometric citation data to derive its rankings. This tendency reflects a global trend by policymakers towards more administratively efficient and ostensibly ‘objective’ measurement processes. However bibliometric data applied inappropriately has the potential to be quite damaging to particular disciplines. Meyer et al., (2009) have noted the tendency to apply inappropriate bibliometrics such as the ISI journal citation databases for the computing disciplines (Meyer et al., 2009), a collection in which as Clear & Young, (2007b) have similarly noted, computing publications are poorly represented.

Features of the Two Schemes
The evaluation methodologies of PBRF and ERA have quite different features, and somewhat differing goals, although the general notion of ranking research performance is common to both.

PBRF
The goals of the PBRF in New Zealand can be viewed from two perspectives, the stated goals and the imputed goals. The official position is outlined below:

“The main aims of the PBRF, as agreed by government, are to:

• Increase the average quality of research
• Ensure that research continues to support degree and postgraduate teaching
• Ensure that funding is available for postgraduate students and new researchers
• Improve the quality of public information on research output
• Prevent undue concentration of funding that would undermine research support for all degrees or prevent access to the system by new researchers
• Underpin the existing research strength in the tertiary education sector”

(TEC, 2010, p.14)

By contrast the imputed rationale has been surmised by Boston (2006) in the statements below:

“In my view, the idea of establishing the PBRF gained significant and broad support across the tertiary sector, especially within the universities, for three pivotal reasons

• The desire to address the inadequate level of funding per EFTS [effective full time student]
• The desire to preserve, if not enhance, institutional and sectoral differentiation
• The desire to enhance TEO’s [tertiary education organisation] accountability, particularly in relation to their research activities”

(Boston, 2006, p.14)

“There was considerable anecdotal evidence that many providers, especially, but not exclusively, those accredited by the New Zealand Qualifications Authority,
were in breach of section 254” [namely that their degrees were not being taught ‘mainly by people engaged in research’] (Boston, 2006 p.21).

“Additionally the PBRF would provide an opportunity – if policy makers chose to use it – to determine which postgraduate programmes were being taught by staff who were not active researchers. On the basis of such evidence, the government would then indicate to the relevant TEO’s that unless appropriate action was taken it would cease to fund such programmes. This would provide an evidence based approach for dealing with the potential problems associated with any proliferation of postgraduate programmes outside the university sector” (Boston, 2006, p. 22).

So the goals of the PBRF can be seen as preserving, if not entrenching, the position of the established universities against the newcomers such as the Institutes of Technology and Polytechnics (ITP) sector. The mechanisms by which the PBRF achieves this differentiation are outlined below:

“The PBRF funding formula is based on three elements or ‘measures’:
• Quality Evaluation: the assessment of the research quality of TEO staff members, based on peer review
• A Postgraduate Research Degree Completions (RDC) measure: the number of postgraduate research-based degrees completed in the TEO
• An External Research Income (ERI) measure: the amount of income for research purposes received by the TEO from external sources” (TEC, 2010, p. 17).

The weightings of these elements are 60% to researcher quality evaluation, 25% for research degree completions and 15% for external research income. The latter two categories by their nature are weighted against non university participants in the scheme, since institutions outside the university sector are not likely to have higher research degrees or significant external research income. The researcher quality evaluations are derived from a six yearly peer review process, in which individual research portfolios from eligible academics (those teaching on degree programmes) are assessed and scores allocated. Each submitted academic’s portfolio is allocated a score by the relevant discipline panel for a combination of the research outputs (RO), contribution to the research environment (CRE) and peer esteem (PE). Scores range from A, B, C (or CNE for a new and emerging researcher), to R or RNE (research inactive for each class of researcher). As an indication the 2006 evaluation results are portrayed in table 1 below.

<table>
<thead>
<tr>
<th>Researcher Quality Score</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>7.42</td>
</tr>
<tr>
<td>B</td>
<td>25.55</td>
</tr>
<tr>
<td>C</td>
<td>24.80</td>
</tr>
<tr>
<td>CNE</td>
<td>9.69</td>
</tr>
<tr>
<td>R</td>
<td>22.08</td>
</tr>
<tr>
<td>RNE</td>
<td>10.46</td>
</tr>
</tbody>
</table>

Table 1: PBRF 2006 Scores: Distribution of Quality Categories (source TEC, 2007, p. 47)

Institutions which have chosen to participate in the PBRF process are then allocated funding annually to support their research on the basis of their aggregate performance against the three funding categories (Researcher Quality, RDC, and ERI).
In the foreword to the ERA report the CEO of the Australian Research Council proudly notes, “ERA is the first comprehensive review of research undertaken in Australian higher education institutions. With the release of this National Report, we are now able to measure our achievements against those of our peers around the world and, in this way, guide our investments into the future” (ARC, 2011).

“ERA... draws together rich information about discipline-specific research activity, at each individual institution, as well as information about contribution to the national landscape of each discipline in each institution” Senator the Hon Kim Carr (ARC, 2011).

The objectives of ERA are to:

1. Establish an evaluation framework that gives government, industry, business and the wider community assurance of excellence of research conducted in Australia’s higher education institutions.

2. Provide a national stocktake of discipline-level areas of research strength and areas where there is opportunity for development conducted in Australia’s higher education institutions.

3. Identify excellence across the full spectrum of research performance.

4. Identify emerging research areas and opportunities for future development.

5. Allow for comparisons of Australia’s research nationally and internationally for all discipline areas.

(ARC, 2011, p.1)

The computing discipline was evaluated under Cluster Five, Mathematical, Information and Computing Sciences and within that cluster was further divided into 08 -Information and Computing Sciences (a two digit field of research 'FoR' code) which consisted of eight further sub categories (four digit codes). Of these further eight only one, Information Systems was a recognisable discipline, with the others being sub disciplines.

The evaluations from each institution were conducted under a discipline based unit of evaluation (UoE). In order to be assessed a unit’s output had meet the threshold of 50 journal articles in the evaluation period. For the ranking exercise conference papers were not considered. This is significant as 75% of the computing disciplines outputs were conference papers. The 2011 report notes “The number of assessed UoEs for Information and Computing Sciences is low relative to the number of research outputs. This reflects the fact that the volume threshold was based on journal articles rather than conference papers” (ARC, 2011, p.85).

Global Research Ranking Processes

As evident from Kim Carr’s earlier comments, countries are concerned with their natural competitiveness. Likewise universities have a global concern with ratings and esteem. “The universities in particular are indeed what economists would call 'prestige maximisers': their bottom line is prestige rather than monetary profit” (Clark, 2004, p.166).

Research recording processes are well known and reported world wide. The OECD Frascati Manual which published its 6th edition in 2002 and claims to have
become “a standard for R&D surveys worldwide” (as cited in Clear 2007) sets out what they believe can be counted in comparing national R&D statistics.

Moed (2006) “highlights important factors that should be taken into account in the interpretation of bibliometric rankings of universities…”. He further goes on to suggest “the creation of a reliable information system on world universities, useful in research management and policy at the institutional, national and supra-national level, and for the wider public” and proposed further steps in this process.

The significant amounts of funding that governments put into Research and Development means that they and the public need some measure of accountability for that funding hence the need for some metrics and ways of assessing performance. Since nations are interested in economic performance and R&D statistics are seen as part of that measure results are sought by politicians and economists.

“The need for policy makers and the wider public to obtain insight into the scholarly quality of research activities in universities is legitimate, but scholarly research quality is not as straightforwardly measured and ranked as performance in many other societal domains.” (Moed, 2006)

As an example of one approach to comparing institutions a profile of a technical university and its research performance at an international level is given in Figure 1.
Figure 2 (Moed, 2006, p. 27)

Such bibliometric exercises based on academic journals, indexed in the Web of Science (WOS) enable consistent performance comparisons and ranking exercises to be conducted at cross-institutional and international levels. However for disciplines with non-journal based publishing cultures (e.g. humanities with monographs as a common form of research output), this form of metric is not valid. Computing suffers from similar issues.

As Meyer et al have observed, "In the computer science publication culture, prestigious conferences are a favorite tool for presenting original research—unlike disciplines where the prestige goes to journals and conferences are for raw initial results. Acceptance rates at selective CS conferences hover between 10% and 20%" (2009).

Echoing this perspective from Europe, the recent conference and journal ranking process undertaken by the Computing Research and Education Association of Australasia (CoRE) has seen the Australian government accept a parity of esteem for selected conferences and journals within the computing and engineering disciplines (CoRE, 2009).

To emphasise the point Meyer et al. (2009) further note: “Journals have their role, often to publish deeper versions of papers already presented at conferences. While many researchers use this opportunity, others have a successful career based largely on conference papers. It is important not to use journals as the only yardsticks for computer scientists”.

The Computing Disciplines- Results Under Each Evaluation

PBRF

The 2006 assessment of the Performance Based research process of that year, rated the individual overall research performance against all other disciplines. Computer Science, Information Technology, and Information Science researchers achieved a 2.75 FTE-weighted quality score just below the average of all disciplines at 2.96 (FTE-weighted).

This is out of a maximum of 10. A score of 2.96 indicates the average quality of the research produced by PBRF eligible staff and indicates this is towards the bottom of the C/CNE range. As mentioned in Chapter 4 of the report "This data has to be interpreted with care" (TEC, 2007).

Despite an earlier gloomy prognosis about the ITP sector impacts “.. for such institutions, teaching workloads will rise to compensate, squeezing out time available for research.” (Clear, 2002) the ITP sector initially fared better than expected, as indicated in the two quotes below.

“In the ITP sector Computer Science, Information Technology and Information Science ranked third as a subject area for proportion of ‘research active’ staff” (Clear & Young, 2007a).

“The NACCQ sector, when considering the proportion of active researchers, outranks the nursing discipline across all sectors, and is not far behind education across all sectors.” (Clear & Young, 2007a)

However since the 2006 PBRF evaluation the funding impacts on the ITP sector have started to be felt more intensively so that earlier gloomy prognosis seems more likely to eventuate.
**ERA**

The summarized outcomes for ICT research performance in 08 Information and Computing Sciences are attached in Appendix 1. This table represents 95% of the computing related outputs in this UoE, with the others being made up of 15 outputs in Commerce, Management, Tourism and Services (1%), 9 outputs in Engineering (1%), and 3% of outputs in the other category. To be evaluated the output threshold was 50 journal articles. The results were then depicted in a six point rating scale.

5 Outstanding performance well above world standard
4 Performance above world standard
3 Average performance at world standard
2 Performance below world standard
1 Performance well below world standard
n/a not assessed due to low volume

As portrayed in appendix 1 a nonsensical ranking exercise was performed by the Australian newspaper, wherein each of the FoR code totals, was averaged to gain an overall score across research disciplines for each university. The results for computing in comparison with this figure even appear very disappointing, however there are a number of mitigating factors. First the already discussed restriction to a threshold of 50 journal articles for each UoE, whereas it has been proven that the majority of ICT outputs are published in conference proceedings (Wainer, Goldstein and Billa, 2011). The citation coverage service that was used by the 2010 ERA was Scopus which does not cover ICT particularly well, and it does not cover conferences (Bailes, 2011).

Only one university, Australian National University (ANU), gained a 5, ‘outstanding performance well above world standard’. This was both at the overall two digit code level and in the sub category Artificial Intelligence and Image Processing with a 3 in the sub category, Computation Theory and Mathematics. 17 or 41% of the universities did not even rate. Of the remaining 24 universities only four, 10%, scored a 4, well above world standard, nine, 22% scored a 3, average performance at world standard, six scored a 2, performance below world standard and a further four scored a 1, well below world standard.

In summary, computing seems to have been reported inconsistently: at one level at the two digit field of research code 08, Information and Computing Sciences, “91% of assessed UoE’s received a rating at or above world standard” (ARC, 2011, p. 85). However in the subsequent institutional comparison tables (ARC, 2011, p. 271, and summarized in Appendix 1) few institutions achieved or surpassed the 50 journal article threshold at the four digit code level. At the two digit code level only 58% of institutions scored at or above world standard. In effect large proportions of computing as a discipline were rendered invisible.

**International Trends**

While the prevailing ‘audit culture’ (Yates, 2005) fashionable with neo-liberal governments has seen instigated such expensive research accountability regimes as the RAE in UK, and the PBRF in New Zealand, these have been fraught, contested and costly to implement. A natural response by treasury ministers then, is to seek a cheaper and ostensibly objective means to achieve their aims. The 2007 statement by the Chancellor of the Exchequer in UK that “for many subjects the RAE based on peer review would be scrapped in favour of the use of bibliometrics to evaluate them” came as a surprise (Oppenheim, 2008).
Subsequent work on how to implement that policy has foundered, and there is now an acceptance that expert peer review panels will still be needed but the use of bibliometric data may supplement their work. For instance the HEFCE consultation response concluded that:
“There was general but cautious support for or acceptance of the use of citation data to inform but not replace expert review in some disciplines” (HEFCE, 2010, p.4).

Complementing the focus on research output assessment the new REF regime in UK (Research Excellence Framework), to be held in 2014 will further assess both the “impact” of the research and the “research environment”, respectively weighted at 65% for outputs, 20% for impact and 15% for environment (HEFCE, 2011, p.2).

The impact element will include all kinds of social, economic and cultural benefits and impacts beyond academia, arising from excellent research, that have occurred during the period 1 January 2008 to 31 July 2013, (ibid. p. 4).

The proposal appears to advocate the use of submitted case studies demonstrating research impact in order to perform such assessments. This UK development is an interesting one especially by comparison with the Australian context where the ERA replaced the previously proposed RQF (Research Quality Framework) model, under which the ‘impact’ of the research was also to be assessed. While determining the impact of research is a laudable endeavour (Clear & Young, 2007b) the difficulties of making such assessments appear to have sunk that controversial scheme, one aim of which was to encourage institutions to “focus better on quality and relevance of their research, that would encourage positive behaviours on the part of researchers” (Yates, 2005). This of course as Yates further noted, was part of a government perspective that viewed research as part of Australia’s global arsenal for economic success in a global market through deepening Australia’s innovation base. However in responses to the RQF consultation process there was “some reservation that it [humanities and social science research] can easily be measured in any short term way” (Yates, 2005).

One assessment of the UK approach is that it results from a conservative government of a similar stamp to the then Howard Government in Australia. But more significantly that the former RAE had run its course in differentiating adequately between the highest and lowest performers, and the old elites were seeking a new vehicle to avoid loss of funding and to reassert their dominance and reputations.

**Implications**
PBRF adopts an evaluation methodology with the individual researcher as the unit of analysis. It is assessed by peer review panels which in the case of computing were sympathetic to the publishing culture and reputational systems. Computing as a discipline fared as “middle of the road” within all academic disciplines. At a sector level the Universities outperformed the ITP sector but the ITP sector itself had made considerable gains in the second evaluation in 2006 (Clear & Young, 2007a).

ERA adopts a bibliometric based evaluation methodology, with a threshold of 50 journal articles (scopus indexed) where the fields of research within an institution are the units of evaluation.

The impact of all of these evaluation processes for computing is threefold (Bailes, 2011).
Funding
First there is the funding threat. If computing is seen as such a low achiever, universities and funding bodies will not consider grant applications from researchers in these areas highly. In Australia access to SRE funding was contingent on participation in the 2010 ERA.

Reputational threats
The reputation of universities and ITPs will be under threat from international student applications as students look to enrol at high performing institutions and the perception among these students could be based on the inconsistent results of research evaluations. A similar situation could well be envisaged among potential PhD students who will be discouraged in pursuing higher qualifications in the computing discipline.

International researchers will be less likely to consider New Zealand and Australia as a potential place to enhance their careers and likewise we may well lose top academics to overseas institutions.

Internal threats
Within institutions there may well be an amalgamation of the computing disciplines thus losing the interdisciplinary aspects of the computing discipline to others e.g. (bio-informatics to biology).

More drastically, institutions that had perceived "poor performance" in the evaluation systems might well decide to close down their computing disciplines as in the quote below:

"More than 20 academics in the computer science department were at risk of losing their jobs, along with a group of world renowned researchers who previously made up the Group of Logic, Language and Computation (GLLC) that spanned the computer science and philosophy departments. The potential devastation resulted from government funding cuts of £1.1 billion in higher education by 2013 and, in the case of KCL’s computer scientists, the poor performance in research rankings of the School of Physical Sciences and Engineering in which they worked" (Underwood, 2011).

Teaching only Academic positions
One response suggests there could be a greater specialization in academic roles. "The proportion of time an average academic spends on teaching is typically equal to or greater than that spent on research. For instance [Ashcroft, 2005] notes that Otago University in New Zealand (an established and traditional PhD granting research intensive institution), "has adopted a generic workload model for its Division of Humanities that recommends that 40 percent of an academic’s time be spent on research and 40 percent on teaching (with the remaining 20 per cent being designated for service to university and community)” (Clear, 2007).

While the quote above provides a breakdown of how an academic’s time is spent in a New Zealand University context, for the ITP sector the balance is more likely to be closer to a 20/60/20 percent split between research, teaching and service. The implications of the ERA on university sector academics in Australia, is hinted at in the excerpt below:

"Proposals at La Trobe University to create a research college of selected academics have been branded divisive by the academic union, which fears it will force some staff into teaching only roles...Members would include only staff judged to be producing research of world class standard...[and] would have an indicative time allocation of at least 40 per cent for research and scholarship. But those excluded would have an
indicative allocation of just 20 per cent for scholarship...NTEU branch president...said heavy teaching commitments kept the average time academics had for research down at 12 per cent to 17 per cent” (Trounson, 2011).

**Subsequent Developments**

Since publication of the ERA results, the whole exercise has become somewhat discredited. In a review of the exercise a number of serious criticisms were raised, among others the following:

“...the Australian Academy of Science argued strongly that key areas such as interdisciplinary research and new research were seriously disadvantaged by journal ranking. This affected not only areas of science and technology, but also interactions between the sciences and the humanities. People whose work is very relevant to Australian issues rather than internationally, and those in new fields or collaborating between several universities, have been particularly disadvantaged” (Creagh, 2011).

As a consequence some significant changes have been implemented:

“The Australian government has dropped the contentious system of ranking academic journals and assessing academics based on their ability to publish in the top-ranked publications...There is clear and consistent evidence that the rankings were being deployed inappropriately within some quarters of the sector, in ways that could produce harmful outcomes, and based on a poor understanding of the actual role of the rankings. One common example was the setting of targets for publication in A and A* journals by institutional research managers” (Creagh, 2011).

In addressing the issues faced by computing CoRE has negotiated with the Australian Research council to conduct a pilot study to investigate whether reliable conference citation data can be accessed from Google Scholar/Microsoft Academic Search (email to CoRE membership from Chair of CoRE, 30/05/2011). Developments on the status of conferences and citation data under the latest revision of ERA await these outcomes.

**Conclusion**

This review of the performance outcomes of the computing disciplines in recent Australian and New Zealand research evaluation exercises, has described the characteristics of the two schemes, compared the relative outcomes for computing and their potential impacts. Under the bibliometrically informed ERA scheme large proportions of computing research endeavour were effectively rendered invisible, whereas under the peer review panel approach of PBRF computing as a discipline fared much better. This suggests that rather than global and necessarily generic bibliometric schemes, discipline conscious schemes have more to offer in delivering equitable outcomes. Therefore while the global context for research performance evaluation schemes suggests the rise of bibliometrics as indicators of performance, the role of expert and peer review cannot be discounted. These international developments need to be carefully considered in the future design of New Zealand’s PBRF evaluations.

**References**


Ashcroft, C. (2005). *Performance Based Research Funding: A Mechanism to Allocate Funds or a Tool For*


Appendix 1

ICT Research Performance scores in 2010 ERA

Two Digit codes

08  Information and Computing Sciences
10  (MIC) Technology

Four Digit codes

0801  Artificial Intelligence and image processing
0802  Computational Theory and Mathematics
0803  Computer Software
0804  Data Format
0805  Distributed Computing
0806  Information Systems
0807  Library and Information Studies
0899  Other Information and Computing Sciences
1005  Communications Technologies
1006  Computer Hardware

Table 1: 2010 ERA results for ICT disciplines (ex Bailes, 2011, P. 4-5)