



United Nations
Educational, Scientific and
Cultural Organization

Organisation
des Nations Unies
pour l'éducation,
la science et la culture

Organización
de las Naciones Unidas
para la Educación,
la Ciencia y la Cultura

Организация
Объединенных Наций по
вопросам образования,
науки и культуры

منظمة الأمم المتحدة
للتربية والعلم والثقافة

联合国教育、
科学及文化组织

APEID, UNESCO Bangkok Occasional Paper Series
Paper No. 2

Performance Indicators for Assessing and Benchmarking Research Capacities in Universities

**V. Lynn Meek
and
Jeannet J. van der Lee**

Centre for Higher Education
Management and Policy
University of New England, Australia

May 2005

TH/2005/PI/H/4

Table of Contents

CONTEXT OF THIS PAPER	1
SCOPE OF THIS PAPER	1
BACKGROUND TO THE ASSESSMENT OF RESEARCH QUALITY	1
PART 1: AN INTRODUCTION TO BENCHMARKING AND PERFORMANCE INDICATORS ..	3
1.1 WHAT IS 'BENCHMARKING'?.....	3
1.1.1 <i>Types of Benchmarks</i>	4
1.1.2 <i>Benchmarking in Practice</i>	5
1.2 WHAT ARE 'PERFORMANCE INDICATORS'	6
1.2.1 <i>Performance Indicators in Practice</i>	6
1.2.2 <i>Applications of Performance Indicators</i>	9
1.2.2.1 PIs and institutional strengthening.....	9
1.2.2.2 PI's and research performance.....	9
1.2.2.3 PIs for funding allocation	11
1.2.2.4 PIs and research infrastructure.....	12
1.2.2.5 PIs for quality assurance.....	12
1.2.3 <i>Dangers of PIs</i>	13
1.3 BENCHMARKING AND PERFORMANCE INDICATORS: A SUMMARY.....	14
PART 2: ASSESSING AND COMPARING QUALITY IN RESEARCH	15
2.1 IMPACT OF RESEARCH	15
2.2 SUSTAINABILITY OF RESEARCH	17
2.3 IMPORTANCE OF RESEARCH.....	17
2.4 POTENTIAL OF RESEARCH.....	18
2.5 INTERNATIONAL COMPARISONS	18
2.5.1 <i>Difficulties of International Comparison</i>	19
2.5.2 <i>Examples of international benchmarking</i>	20
2.6 CRITERIA OF QUALITY IN RESEARCH.....	21
2.6.1 <i>Measurements of Quality</i>	22
2.6.1.1 Subjective evaluation: peer-review.....	22
2.6.1.2 Number of Publications	23
2.6.1.3 Research publications index	24
2.6.1.4 The number and value of research grants and contracts gained.....	25
2.6.1.5 Quantitative vs Qualitative Indices.....	25
2.7 EXAMPLES OF RESEARCH QUALITY ASSESSMENT PROCESSES AND THE INDICATORS USED.....	27
2.7.1 <i>Higher Education Research Data Collection - Australia</i>	27
2.7.2 <i>Australian National University Research Quality Review 2004</i>	27
2.7.3 <i>The Research Assessment Exercise – United Kingdom</i>	28
2.7.4 <i>League Table Approach</i>	28
PART 3: DESIGNING A RESEARCH EVALUATION PROCESS	29
3.1 PURPOSE OF REVIEW	30
3.2 ALIGNMENT WITH MISSION	30
3.3 REGIONAL RELEVANCE	30
3.3.1 <i>How can it be measured?</i>	31
3.4 SCALE.....	31
3.5 INDICATOR SELECTION	31
3.5.1 <i>Inter-university comparison and data standardisation</i>	32
3.6 ACTION PLAN	33
PART 4: CONCLUSION	34
REFERENCES	36
APPENDIX 1: THE MCKINNON BENCHMARKS.....	41
APPENDIX 2: DETAIL OF RESEARCH ASSESSMENT EXAMPLES	43
A2.1 Higher Education Research Data Collection - Australia	43
A2.2 Australian National University Research Quality Review 2004	44
APPENDIX 3: THE CHARLES AND BENNEWORTH BENCHMARKS.....	46

Context of this paper

At the first GUNI-AP (Global University Network for Innovation – Asia and the Pacific) symposium in 2004 on Quality Assurance and Mutual Recognition of Qualifications in Higher Education, it was agreed that the theme for the 2005 Symposium would be *University Capacity Building and its Benchmarking*, with the goal of a regional benchmarking policy for university research capacity building. This paper is the first step toward that goal, through providing professional input into the performance indicators for assessing/benchmarking research capacity in universities.

Specifically, this paper was commissioned for the dual purposes of:

1. assessing the quality of research programmes in a university to identify strong/good quality programmes for priority funding. The criteria of quality were to include:
 - a. impact of research
 - b. sustainability of research
 - c. importance of research
 - d. potential of research
2. enabling a university to benchmark its research capacity with other universities found in the country/region for the purpose of ranking/rating the universities in terms of their research capacity/performance. Performance indicators to be developed for assessing different levels:
 - a. fields of study/research
 - b. research programs
 - c. research institutions

This will be achieved through extensive literature review regarding the “state-of-the-art” in research performance benchmarking and the indicators that are commonly applied, along with the processes that are usually employed in achieving effective and meaningful benchmarking.

Scope of this paper

This paper seeks to present a review of the current practice in benchmarking and to outline the types of performance indicators that are used, particularly in relation to research within Universities. The paper commences in Part 1 with a general overview of benchmarking and performance indicators, particularly in relation to their definition, application and processes for their use. Next, Part 2 provides some detail about assessing and comparing the impact, sustainability, importance and potential of research and outlines the types of indicators that might be used to achieve this. Recommendations are then made in Part 3 as to the elements that need to be considered in designing a research evaluation process. Part 4 concludes with a few summary comments. This paper does not specify a particular procedure for benchmarking or prescribe specific indicators for use by the Asia-Pacific Universities. Rather a starting point from which this group can develop their own benchmarks and measures that are appropriate to their context are presented.

Background to the Assessment of Research Quality

Interest in benchmarking and the associated use of performance indicators is not an activity restricted to the higher education sector. It is more that higher education is caught in a broad web of socio-economic and political imperatives designed to make all public sector institutions more cost efficient and internationally competitive. Performance indicators

are proxies for many things, not the least of which in the public sector is the harsh reality of market competition.

The current use of performance indicators in many countries is being driven by the desire on the part of government to introduce more market-like competition into higher education in order to make institutions more flexible, cost-efficient and responsive to the needs of society. But, somewhat ironically, higher education policies based on market principles may be having an effect opposite to that intended. There is growing evidence that market competition in higher education may stimulate bureaucratic rigidity rather than flexibility and diversity (Marginson 1993; McElwee 1992; Meek 1995). One must be careful that the development and implementation of performance measures for the purpose of benchmarking in higher education does not undermine the very responsiveness and quality that they may be intended to enhance.

PART 1: An Introduction to Benchmarking and Performance Indicators

Garlick and Pryor (2004:28) found in their research that “[t]here is considerable uncertainty as to what benchmarking is really all about and what its benefits for the organisation and its staff and students might be. There is confusion between benchmarking, quality auditing and the need for quantitative key performance indicators”. It will therefore be useful to initially provide some definitions and examples to illustrate the differences.

1.1 What is ‘Benchmarking’?

Benchmarking has been defined by Jackson and Lund (2000:6) as:

Benchmarking is, first and foremost, a learning process structured so as to enable those engaging in the process to compare their services/activities/ products in order to identify their comparative strengths and weaknesses as a basis for self improvement and/or self-regulation.

Garlick and Pryor (2004:19) have built upon this notion in their work to further characterise benchmarking

in the university situation [as using] terms like collaboration, organisation learning, inclusiveness, reflection, review, leadership and improvement. This way is about connecting up relevant stakeholders both within and outside the institution in such a way that leads to knowledge exchange about why, what, where and how improvement might occur.

It follows then, that when undertaking a benchmarking exercise it will be imperative:

1. that there is a strong commitment (including financial) by the institutional leaders to reflect upon the outcomes and undertake improvement where required; and
2. that a collaborative process is initiated and commitment is assured by participating organisations and stakeholder groups.

In relation to the last point, Garlick and Pryor (2004:3) found in their research that there is a tendency to apply the measures as a reactive performance accounting exercise, rather than as a process involving participation, reflective dialogue and learning with the aim of identifying areas for improvement. It has also been noted by Martin (2003) that institutions have found it difficult to “close the loop” and incorporate the feedback that benchmarking provides into the improvement of their operations. We would argue here, that this reflexivity is the most important function of the benchmarking process, as it is the feedback that will inform continuous improvement and enable institutional strengthening.

Benchmarking has become a mainstay in the process of continuous improvement (Charles & Benneworth 2002). In relation to this, Garlick and Pryor (2004:9) further described benchmarking as having two objectives:

... first, as a means for assessing the quality and cost performance of an organisation’s practices and processes in the context of industry-wide or function-specific ‘best practice’ comparisons. This has been generally used as part of an organisation’s accountability responsibility to an accrediting, funding or regulatory authority.

Second, and more fundamentally, benchmarking can be used as an ongoing diagnostic management tool focused on learning, collaboration and leadership to achieve continuous improvement in the organisation over time.

Furthermore, “[i]t is important that benchmarking as a process is owned by all staff and does not just remain a solo activity undertaken by a specialist unit within the university or be the responsibility of senior management” (Garlick & Pryor 2004:29).

According to Charles and Benneworth (2002:4)

several aspects of benchmarking [that] are valuable [and are worth noting here include] :

- a. A range of benchmarks can cover the variety of actions undertaken by a heterogeneous set of HEIs [Higher Education Institutions]. Any HEI can identify at least some areas in which it can be successful, rather than being assessed against a few externally selected criteria.*
- b. Benchmarking can support decision-making on where HEIs should devote more effort to support their regions more effectively. This can be usefully connected with a parallel exercise at regional level to determine regional priorities.*
- c. Benchmarking approaches have been developed which use qualitative as well as quantitative indicators, process measures, and leading and lagging indicators. They can therefore help to identify if good practices are being adopted without waiting to measure outputs.*
- d. Benchmarking allows the combination of different forms of measurement, and models of performance.*

1.1.1 Types of Benchmarks

McKinnon, Walker and Davis (2000:7) provide a useful summary of the main approaches to the formulation of benchmarks. They distinguish *criterion reference* and *quantitative* as the two types of benchmarking approaches. Where the criterion reference approach defines the attributes of good practice in a particular area, thus enabling universities to benchmark their success in that area through a direct comparison of their performance against the criterion. In contrast, quantitative benchmarks distinguish normative and competitive levels of achievement, enabling assessments to be made of differences in practice between institutions. This “numbers focus” has been criticised by Garlick and Pryor (2004:19) because “...[i]t gives emphasis to performance assessment rather than improvement, and it segments rather than joins up functional areas, thereby limiting the learning process about where improvement can occur and limiting longer-term commitment to implementation”.

McKinnon, Walker and Davis (2000:7-9) also outline the key issues that need to be considered in intra and inter-university benchmarking exercises. These include:

- *a focus on outcomes rather than process,*
- *a focus on good vs best practice, where good practice is preferred and best practice is defined by the highest level of practice identified in the benchmark;*
- *testing for continuous improvement;*
- *identification of benchmarks to measure functional effectiveness rather than countables;*
- *adjusting for inequalities in university features, so that benchmarks can be expressed as proportions, times, and ratios.*

- *the choice and use of benchmarks should reflect the diversity of multi-campus universities;*
- *specification of scale for benchmark use, eg. University, faculty, school/department;*
- *re-calibration of benchmarks to reflect improved data definition and collections;*
- *choosing benchmarks that are valid and convincing;*
- *organisation of benchmarks into functional groups;*
- *expert judgement to achieve objectivity; and*
- *the number of benchmarks need to be manageable, as they will need to be monitored regularly.*

1.1.2 Benchmarking in Practice

Benchmarking in Australian universities is a relatively recent practice but has to date been largely restricted to an assessment of administrative function (eg. libraries, laboratories, facilities etc.) rather than focusing on teaching and research. The dominant attitude toward benchmarking seems to be one that seeks to maximise reputation, rather than as a reflective exercise in institutional strengthening. In relation to this, Weeks (2000 in Garlick & Pryor 2004:13) “concludes that when used as a process for generating ongoing improvement, as opposed to it being an indicator of competitiveness, benchmarking can be effective in influencing culture and practice within the university through internal learning, and in building networks with professional colleagues in related fields”. Having said that though, a growing trend in benchmarking the relevance of the university in its regional community has been noted as a recent occurrence (Garlick & Pryor 2004:9).

Garlick and Pryor (2004:vii) found that benchmarking in Australian Universities has become more widespread over the past 15 years, with some further stimulus provided by the release of such documents as the McKinnon manual in 2000, *Benchmarking: A Manual for Australian Universities*¹ (McKinnon, Walker, & Davis 2000). This manual provided tools for performance assessment to feed into continuous improvement activities and proposed a set of 67 indicators² (9 relating directly to research) with which to do so. However, Garlick and Pryor’s (2004: vii) research further suggests that the use of benchmarking “...has, in the main, been superficial and peripheral. It has generally not been used for organizational improvement in core business areas. Its development has not kept pace with changes in recent years in performance reporting and quality auditing...”.

One of the main problems with benchmarking between universities has been a lack of consistency in the benchmarks used and the method of measurement, rendering comparison very difficult.

¹ Garlick and Pryor (2004:viii) “found that the McKinnon et al. manual was not seen positively as a benchmarking tool that could assist universities with their improvement agenda. It was seen as having confusing language and concepts, as being a ‘one size fits all’ top– down approach, and as anathema to ‘learning for improvement organisations’. It was seen to also contribute further to the existing uncertainty and suspicion with which evaluation and auditing processes generally are regarded in universities. It is argued that organisational improvement is a more personal process for staff and stakeholders than that fostered by a template-based manual of the type put forward by McKinnon et al. It needs to incorporate collaboration and connectivity across stakeholder interests, learning and knowledge exchange, and leadership commitment. And it needs to be simple. We therefore do not advocate any updating of the approach to university benchmarking advocated by McKinnon et al. in the current manual. It does not offer a solution to benchmarking in an increasingly ‘involve me’ evaluation world.”

² These have been reproduced here as Appendix 1.

1.2 What are ‘Performance Indicators’

Goedegebuure, Maassen and Westerheijen (1990:29) note that “an authoritative definition of performance indicators does not exist”, and while they are often thought to be solely quantitative, “qualitative performance indicators exist too”. For example, Cuenin (1987) acknowledged that at the very least a performance indicator is a numerical value:

which provide[s] a measurement for assessing the quantitative performance of a system. When the indicator shows a difference in one direction, this means that the situation is better whereas if it shows a difference in the opposite direction then this means that the situation is less favourable.

Dochy and Segers (1990 in Ball & Wilkinson 1994:418) however extend their description into the realm of qualitative data, by stating that:

A first requirement is that they should be clearly related to the defined functions of the institution. A second requirement is that they are only what their name states, indicators of the extent to which institutional goals are achieved. A third requirement is that they should be a valid operationalization of what they intend to indicate and that they can be measured and interpreted in a reliable and correct way.

Cuenin (1987) is one of many commentators who has recognised the need for greater clarity in definition of performance indicators, a challenge taken up by various peak bodies such as the Organisation for Economic Co-operation and Development’s (OECD) Institutional Management in Higher Education’s Performance Indicators Workgroup initiated in 1989. The Workgroup defines performance indicators as “signals derived from data bases or from opinion data that indicate the need to explore deviation from either normative or other preselected levels of activity or performance”; indicators “monitor developments or performance, signal the need for further examination of issues or conditions, or assist in assessing quality” (Kells 1992:133). Kells (1992:133) also pointed out that there are at least three different types or categories of indicators: “1) Indicators to monitor institutional response to government goals or policies; 2) Indicators of teaching/learning, research and service; and 3) Indicators needed in university management”.

1.2.1 Performance Indicators in Practice

The heightened interest in performance indicators is due to many factors, not the least of which are an institutional desire to implement more effective strategic planning mechanisms and government demands for greater efficiency and effectiveness on the one hand, and enhancement of the quality of higher education on the other. The increased use of performance indicators needs to be seen as part and parcel of a broader move to entrench an 'evaluative culture' (see Neave 1988) in higher education institutions and systems, fuelled by the dual demands of accountability and quality assurance.

Moreover, performance indicators are a management and/or policy tool – they are an aid to good judgement and not a substitute for it, a point made time and again in the literature. Cave, Kogan and Hanney (1990:47) argue that “the extension of performance measurement is inevitable and will have a major effect on the internal ethos of higher education institutions. It is thus vital to be aware of the potential dangers as well as benefits of performance indicators”.

The use of indicators in measuring organisational performance in higher education has moved through a number of distinct phases (see Doyle 1995). The first being their use as

a mechanism to allocate resources in the late 1960s; moving through their use as a source of information input into funding decisions and assessing achievement of national priorities in the 1980s; to a distinct shift in their use in the 1990s for quality assurance and control.

Sizer, Spee and Bormans (1992:137) list five core uses of performance indicators: monitoring, evaluation, dialogue, rationalisation and resource allocation. Harris (1993:22) in attempting to arrive at a definition of performance indicators identifies what he sees as five important aspects:

1. *They are expressed numerically;*
2. *They relate inputs to outputs (ie they measure efficiency of resource use);*
3. *They are linked to the overall goals of the organisation (ie they are concerned with effectiveness in meeting desired outcomes);*
4. *They allow users to determine how the performance of the individual or organisation under study has changed over time and/or how it compares with the performance of other individuals or organisations;*
5. *They may be used as incentives to influence the activities in socially desired ways.*

Several of the above aspects identified by Harris are controversial, particularly with respect to whether performance indicators are only qualitative, the relationship between inputs and outputs as an efficiency measure and the use of performance measures as an incentive device anyway tied to funding.

There is thus a growing literature on both the use and abuse of performance indicators in higher education (see, for example, Cave, Hanney, & Kogan 1991; Doyle 1995; Goedegebuure, Maassen, & Westerheijen 1990; Johnes 1988; Johnes & Taylor 1990; Kells 1990; 1992; 1993; Linke 1992; Politt 1990; Sizer 1992; Stolte-Heiskanen 1992; Yorke 1995). While some commentators appear to imply that the application of performance indicators to higher education is increasingly becoming an 'exact science' with its own specific definitions and methodologies, others strongly disagree (for example, see Ashworth & Harvey 1994; and Cave, Hanney, & Kogan 1991; or Politt 1987; and Sizer 1988).

Performance indicators are often used for one of three purposes: improvement, planning or accountability. The three purposes are not mutually exclusive, but they are driven by different forces. Improvement and planning are usually questions of concern to individual institutions, and it can be argued that "performance indicators are most valuable within the specific institutional context" (Findlay 1990:125). Questions of accountability, however, are raised mainly by governments, which in turn places the use of performance indicators squarely in the political arena, sponsoring wide-ranging debate and vehement criticism. According to Findlay (1990:125), "the mounting pressures for greater accountability, greater value for money, and latterly for evidence of quality of delivery, have converged with the broad debate about the definition of performance indicators". Moreover, the continued interest in performance indicators has been mainly government driven (Kells 1993; Yorke 1995).

Dochy and his colleagues (1990a:71) draw a clear distinction between management statistics (quantitative data), management information (data which are related to each other and structured as management information) and performance indicators which are:

empirical data – quantitative and qualitative – which describe the functioning of an institution, the way the institution pursues its goals. This implies that they are

context- and time-related. Indicators are more general signals which can be translated in a combination of more specific characteristics, called variables.

Another question often central in discussions of performance indicators is their core uses. Dochy, Segers and Wijnen (1990a:71) identify four core uses: evaluation, monitoring, planning and dialogue:

Evaluation concerns the possibility to comment on the degree of goal-attainment. Monitoring refers to the possibility to register developments in the system. Performance indicators can also play a role in the accomplishment of a coherent policy making process. Finally, they are important for the improvement of the mutual administrative relationships [between institutions and government].

A review of the literature indicates that there is a good deal of debate and controversy over the use of performance indicators. But one can detect some general areas of agreement. First, “one must not be tempted to use a multitude of evaluative data. It is necessary to use a limited number of indicators with high validity ...” (Dochy, Segers, & Wijnen 1990b:136). Collecting data for its own sake is a meaningless task.

Second, Jongbloed and Westerheijden (1994:48) note that “doubts about the validity of what is measurable, especially if the object is quality, have led to some disenchantment with PIs. More than ever before, it has become apparent that numerical indicators are signals that require careful interpretation before they are used in decision making”.

Third, some commentators have questioned the validity of applying an input-throughput-output analogy drawn from manufacturing to higher education. Nonetheless, it is quite apparent that many governments are or intend to apply such a model to the funding of higher education, at least in part.

Fourth, performance indicators are an aid to good judgement, not a substitute for it.

Fifth, performance indicators can be both quantitative and qualitative, and it is a somewhat futile task to draw any absolute distinction between performance indicators and the more subjective process of peer judgement. In fact, many numerical indicators, such as publication rate, are ultimately based on peer evaluation.

Sixth, a statistic or measure becomes a performance indicator only when it is infused with value and worth and used in assessing whether specific objects are being achieved – “appraisal and accountability become meaningful only in terms of declared and agreed objectives” (Elton 1988:207).

Finally, it appears that the use of performance indicators is being subsumed in the broader process of quality assurance.

The literature distinguishes several different types of performance indicators, such as input, process and output indicators (Cave et al. 1998). A further distinction can be made between indicators of economy (adhering to budgets), efficiency (ie productivity, or output per unit of input), and effectiveness (degree of attainment of objectives) (Cave, Kogan, & Hanney 1990:48). They can also be considered in terms of relevance in relation to a department's or institution's objectives and mission; degree of ambiguity and ease of interpretation; cost of collection; level of aggregation; uniqueness (Cave et al. 1988); dependability; and durability (Linke 1991). The purposes to which performance indicators can be put are related, in part, to whether they are being applied at the program, institution or system level (Doyle 1995; Jongbloed & Westerheijden 1994).

Dochy, Segers and Wijnen (1990b:139-140) list four essential criteria to select performance indicators:

A first requirement is that they should be clearly related to the defined functions of the institution. A second requirement is the recognition that they are only what their names state, indicators of the extent to which the institutional goals are achieved. A third requirement is that they should be valid operationalisations of what they intend to indicate and that they can be measured and interpreted in a reliable and correct way. Finally, given that their values are relative rather than absolute, their incorporation in the machinery of "management technology" should be approached with extreme caution.

1.2.2 Applications of Performance Indicators

There are a number of settings in which indicators of performance are used to evaluate institutional performance both within and between universities. The following outlines a number of these, including indicators for research performance which is the focus of this paper. However it is important that evaluation of research performance is contexted within the broader institutional evaluation process.

1.2.2.1 *PIs and institutional strengthening*

Perhaps one of the most useful outcomes of a benchmarking exercise is the potential institutional strengthening that can result from a strategic review of the university's performance and overall market placement. In order for this process to be useful however, the indicators that are used must be aligned with the mission, there must be commitment from university management to reflect upon the outcomes of the review and to implement strategic changes as required.

This aspect of course has application in assessing research performance because the benchmarking process will enable reflection to take place within those parts of the institution that undertake research and thus enable them to reflect upon their successes and possibly design solutions to strengthen those areas in which they are found to be under performing.

1.2.2.2 *PI's and research performance*

Assessment of research performance, particularly for funding purposes, is a complex and detailed topic, which can only be dealt with here in a summary fashion. But, for various reasons, the academic community may be more tolerant of assessment of research performance than of teaching performance. This may be due to the fact that external performance measures, usually in the form of peer review, are a routine activity of the research enterprise. Scientists subject themselves to assessment every time they submit an article to a refereed journal or apply for support from an external research funding agency.

The application of performance indicators to the measurement of research productivity seems to be less problematic than their application to teaching/learning activities. Also, numerical measures (such as number of publications or income from research grants per staff member) appear to lend themselves much more to the research enterprise. But the application of performance indicators to research is not as straightforward or non-controversial as it may seem.

Much of the judgement of quality and worth of scientific research has been subjective and largely left to the science experts themselves. But, over the last few years, there has been a growing trend amongst a number of governments and governmental policymakers to adopt more 'objective' criteria and externally imposed input-output measures in order to more rigorously and empirically assess scientific performance, particularly for funding purposes (Wood & Meek 1993).

Ronayne (1984:96) argues that “if a rational, objective approach to science policy is possible, then decision-makers must be able to assess the costs and benefits of particular courses of action in a quantitative way”. But no science/research policy system is ever completely rational; the business of science is too messy, incremental, *ad hoc* and subjective to be entirely rational. Nonetheless, policymakers require hard data wherever possible, and this is certainly true with respect to funding scientific research as cost spirals nearly everywhere. But, as is the case with all types of performance indicators, the danger of an overly simplistic application of measures and techniques to evaluating research outcomes is that data become a substitute for decision-making (Linke 1992).

What is at issue is not assessment of research quality, but the need for policymakers to know that they are receiving value for money. But it is important to note that, with respect to evaluating research for funding purposes, more actors are involved than just university researchers and government bureaucrats.

The ability to attract research funds depends upon the research focus of the institution (whether that be at the university, faculty or school level) and whether or not these are aligned with National Research Priorities. As an example, the current National Research Priorities in Australia are: i. An Environmentally Sustainable Australia; ii. Promoting and Maintaining Good Health; iii. Frontier Technologies for Building and Transforming Australian Industries; and iv. Safeguarding Australia (ARC 2003:2). The Australian Research Council has committed 34% of the total funds available under their programs to these priorities in the funding rounds of 2003-2007 (ARC 2003:4). The ability of an institution to win research funds will therefore depend somewhat upon their capacity to fit within these priorities, and if they do not should they also be penalised as a consequence of performance indicators that are biased toward institutions that do? Simply because an organisation's research focus does not fit into a current research priority does not automatically correlate with institutions whose research is unworthy. It is important then to also consider performance indicators that are able to account for the impact of their research. However, this area is also fraught with difficulties.

While research performance is usually measured and compared first of all at the departmental level, there are few if any absolute standards by which to measure research performance at this level. A department's research performance can only be judged relative to that of similar departments operating under similar conditions elsewhere (Hattie 1990; Linke 1991), and even here such comparisons are problematic for even if departments cover exactly the same disciplinary mix, it is nearly impossible to adjust for historical and institutional contextual factors (Grigg & Sheehan 1989). Bourke and Martin (1992) in their bibliometric³ study of university research performance in the UK and Australia have produced evidence to question “whether the department is actually the most appropriate unit of analysis for evaluating university research”.

3 Bibliometrics are described later in this paper.

Harris (1993:22) notes that it is important “to distinguish between four related but distinct concepts used in evaluating research – quantity, impact, quality, and importance. It is easy and straightforward to measure quantity numerically, and number of citations is often used as a surrogate measure of impact of a piece of research. “Quality and importance, however, cannot be objectively measured: the assessment of quality is highly dependent on value judgements, and the importance of a piece of research may not become clear until time has passed” (Harris 1993:22).

Harris (1993:24) in his review of performance indicators in assessing research argues that 'underneath it all, we are principally concerned with issues of quality, impact and outcome rather than simply performance.

Findlay (1990) also makes the point that when indicators are imposed top-down and/or externally, there will often be much resistance at the 'grass roots' level. It is better that indicators are negotiated between those who require their use for managerial purposes and those whose performance is to be assessed, particularly if quality improvement is the aim.

There is as much if not more debate and controversy over the quality assurance movement as there is about use of performance indicators per se. It is not the intention here to review that debate, but to make the point that views on performance indicators differ depending on whether they are used primarily for accountability and managerialist purposes or for quality improvement.

1.2.2.3 PIs for funding allocation

The use of performance indicators, particularly in relation to research performance, to inform the allocation of block funds to universities has become common practice internationally. This is where performance indicators become most controversial as the emphasis shifts from their use as one of many inputs into effective decision-making to using them as a ranking device to differentially allocate esteem and funding.

Several commentators believe it is detrimental to closely tie performance measures to funding, particularly with respect to quality. Linke (1992:202) cautions that: “There are serious risks ... in applying a performance based funding system which at any particular time has more than marginal influence on institutional recurrent budgets”.

In reality, a small proportion of lecturers produce most of the research, and many of these are little interested in teaching. Performance-based funding of research simply recognizes these facts.

Somewhat paradoxically in an age of mass higher education, research performance is possibly the most important factor for assessing the standing of the modern university. The use of selective performance-based measures of research has served to pull into sharp focus the policy issues of whether differences between institutions and types of institutions are important, of whether such differences have increased or declined, of whether research should be further concentrated in 'research universities', and the nature of the opportunities presented by a more competitive market for universities to enhance their reputation by increasing their research output. (Ramsden 1999:342).

“A related issue raised by performance-based funding of research concerns the predictors of research performance. The topic has been widely studied at the level of the individual faculty member and department....., but much less so at the level of the university” (Ramsden 1999:343).

1.2.2.4 PIs and research infrastructure

Where performance indicators appear to play an increasingly important role is with respect to funding of research infrastructure, which is one reason why there is so much activity in this area. Australia has adopted a policy of selectivity and concentration with respect to research funding and, though the bulk of the financial resources for research are assumed to be part of the general operating grants⁴, a proportion of research infrastructure is differentially funded on the basis of performance.

Currently, the competitive allocation of research infrastructure funding in Australia occurs through three schemes: the Institutional Grants Scheme (IGS), Research Training Scheme (RTS), and Research Infrastructure Block Grants (RIBG) (for further information on these funding schemes see DEST 2004a; b; c). The RTS is based on the number of research student completions (weighted at 50%), research income (weighted at 40%), and research publications (weighted at 10%) (DEST 2004a:7). The IGS incorporates the previous research quantum and small grants schemes and is allocated on a performance based formula consisting of research income (weighted at 60%, Commonwealth-funded research student load (weighted at 30%), and research publications (weighted at 10%) (DEST 2004b:4). The RIBG is allocated according to an institution's success in obtaining Australian Competitive Research Grants (DEST 2004c:5). In 2003, the three schemes combined accounted for well over \$1B in total university funding or about 11% of total funding (DEST 2005a).

The competitive allocation of research funds has made Australian universities highly conscious of the importance of their research profile, some would say to the detriment of teaching. Sizer (1990b:22) recognises that "one of the dangers of selectively supporting research but not teaching is that greater emphasis is placed on research at the expense of teaching".

1.2.2.5 PIs for quality assurance

It is possible (and valuable as well) to assess the degree to which a higher education institution is achieving its stated mission, goals and aims, and also it is possible to assess the relevance of the institution to the needs of society. But if this is to occur, both the goals of the higher education institution and the needs of society must be pre-specified in such a way that lend themselves to assessment. Also, there should be some correspondence between the goals of the institution and their relevance to society, otherwise it is possible to end up with quality institutions of little relevance, or vice versa. However, in that there are a variety of stakeholders with diverse and sometimes divergent views and interests involved in determining both the goals of the higher educational institutions and the needs of society, specification of goals and needs becomes problematic indeed.

⁴ It has been assumed that all academic staff in Australian universities engage in teaching and research and that funding reflects this with approximately a 75%/25% division in funding between teaching and research. However, this assumption is presently under challenge by the federal government.

Until recently, the more informed view was that performance indicators was a device or tool to aid planning and decision-making. But there is some evidence to suggest that at least in some countries performance indicators are becoming part of the process, particularly with respect to quality assurance.

1.2.3 Dangers of PIs

It is apparent that, broadly speaking, the use of performance indicators can be useful indeed when used thoughtfully and carefully. However Bird et al. (2005:1) argue in relation to the public service that when “done badly, [performance monitoring] can be very costly and not merely ineffective but harmful and indeed destructive”. The reasons for this are related primarily to process rather than application, including political interference, lack of protocol, unrealistic aspirational or motivational targets that demoralise and little consideration of perverse impacts (see Bird et al. 2005 for further details).

An Australian report (AVCC/ACDP Working Party 1988:1) cautions that:

Something resembling a Cargo Cult seems to have grown up around the notion of performance indicators, so that all manner of powers and virtues are ascribed to them and expectations are aroused that by collecting and using them great benefits will miraculously result. This erroneous attitude arises when performance indicators are considered in isolation and out of the context in which they should properly be used ... It is much more important to define the different purposes for which indicators can be of assistance than it is to concentrate on the indicators themselves.

In relation to this, Ball and Wilkinson (1994) argue that a particular worry relates to the uncritical interpretation of performance measures and that the use of performance indicators to compare organisations will in actual fact inhibit diversity. They further argue that “the danger of a national system of performance indicators ... is that it will inevitably tend to “nationalise” institutional policy making” (Ball & Wilkinson 1994:426).

Moreover, performance indicators are a management and/or policy tool – they are an aid to good judgement and not a substitute for it, a point made time and again in the literature and in this report as well. Cave, Kogan and Hanney (1990:47) argue that “the extension of performance measurement is inevitable and will have a major effect on the internal ethos of higher education institutions. It is thus vital to be aware of the potential dangers as well as benefits of performance indicators”. Once a management statistic is infused with values and becomes a performance indicator, it loses neutrality: “Even when cloaked in the credibility conferred on them by their mathematical configuration and their seemingly sophisticated formulation, performance indicators are not ‘neutral’, any more than the technical-scientific epistemology that underlies and often shapes them” (Lucier 1992:213).

Doyle (1995:4-5) provides a useful summary of assumptions and limitations to approaches to performance measurement in higher education:

1. Quantitative and qualitative performance data need to be collected by institutions for internal planning and external reporting purposes ...
2. The performance thrust in higher education raises complex and controversial issues relating to the nature, collection, interpretation and use of performance data ...
3. Performance information (performance indicators) are best related to goals of activities, best interpreted by persons or groups responsible for activities, and

best regarded as signals to monitor developments or performance, the need for examination of issues, or to assist in the assessment of quality ...

4. A major problem in the development of performance information is the availability of adequate and reliable data, most of which relates to educational inputs rather than outputs. As information systems improve, there is also a danger in increasing complexity of measures which generates resistance in institutions ...
5. Another major problem is the comparability of measures. Frequently organisational units are not similar in kind and are subject to considerable variability, thus reducing the reliability of comparative data.

Tognolini, Adams and Hattie (1994:116) also raise some important points.

Performance indicators may lead managers and the government to become less internationally orientated; it is not obvious that performance indicators will necessarily enhance and stimulate research; [...] If they are used to maximise research output alone, without attention to the culture, ethos and inputs of university departments, they will probably aim to minimise variability in research performance which would be counter productive.

There is no one clear trend in the literature on how best to use performance indicators in higher education. But there is a degree of general agreement that they are of limited value, must be approached with caution and are effective only when set against the achievement of pre-specified institutional goals. This would infer then, that comparison between institutions would become difficult as they will presumably formulate different indicators that relate to their unique context and organisational goals.

1.3 Benchmarking and Performance Indicators: A Summary

From the preceding discussion it is apparent that benchmarking and performance indicators are two distinct processes that should, in practice, operate in parallel. This means that benchmarking without appropriate performance indicators is a futile exercise. The converse is also true, that is, performance indicators presented outside a benchmarking process are essentially meaningless as they have no context, nor infused with the power to initiate change for improvement in an organisation. According to Tognolini, Adams and Hattie (1994:116) “[i]f used to monitor trends and progress towards some specified objectives, then performance indicators could become developmental rather than punitive. The department would at least know the rules and goals to strive towards”. As such it is imperative when a benchmarking exercise is to be undertaken that there is commitment by the management to incorporate the feedback provided by the performance indicators (that have been selected to reflect the organisational context) into a process of continuous improvement.

PART 2: Assessing and Comparing Quality in Research

The first stated purpose for this research paper related to the assessment *of the quality of research programmes in a university to identify strong/good quality programs for priority funding*. Furthermore, the criteria of quality were to include:

- a. impact of research
- b. sustainability of research
- c. importance of research
- d. potential of research

The following discussion will therefore loosely conform to this structure.

Concentration and selectivity have become the global norms of research funding, due to both the sheer cost of many research areas and the desire of institutions and governments alike to ‘pick the winners’. No nation can afford to fund all of its higher education institutions at the level of world-class research universities and even world-class research universities cannot afford to support all areas of science. That said, however, it needs to be recognised that through identifying only strong/good programs for priority funding within a university that this may seriously undermine the growth of research within that organisation. This would occur because up and coming areas of research would not be identified and thus not allocated sufficient funding to start-up. Similarly, research areas within universities that are ‘flagging’ may need an injection of funds to reinvigorate their research effort, or perhaps some other kind of support that a benchmarking exercise may help to identify. We would advocate here that the use of performance indicators for funding allocation is not appropriate unless it is imbedded within the broader context of an institution’s overall research mission and objectives and that mitigating circumstances with respect to ‘under’ performance can be identified and dealt with.

2.1 Impact of Research

In this paper we will define the ‘impact of research’ as the effect that research had – has - or - will have upon the broader environmental, economic, socio-cultural context. Whether this impact occurs in the short-term or the long-term is also an important factor to consider, although it is obviously simpler to account for the immediate impact of the research because the longer term impact of the research is often not known due to gaps in current understandings. It is often noted that Rutherford, for example, believed that his line of research would never have any practical application. Furthermore, the impact of research may be felt in a number of sectors in different directions. For example, the research outcome might have a positive impact upon the economy in the short-term but a negative effect upon the environment in the longer term and as such reconciling these contradicting outcomes into a measure of impact becomes difficult. Sornette and Zajdenweber (1999) stress that seeking to determine the economic returns from research over a particular period is highly unreliable. It is therefore a considerable challenge to determine relevant indicators of research impact, it is also important that research take place to determine the processes through which impacts are generated by research (Allen Consulting Group 2005).

There is at present a push by governments internationally (see Slaughter & Rhodes 2003) for research to bring immediate short-term economic returns, however what is not considered is that such a policy potentially has long term dangers because the possible future impacts are not considered.

More direct types of impact, and those that are most commonly measured, include those upon the wider research community in relation to the communication and dispersal of the research. When impacts are measured outside the academic/research community it is the economic impact (including “direct commercialisation of publicly funded research, adoption of knowledge by industry and through skills formation improving labour productivity”) that receives most attention (Allen Consulting Group 2005:vii).

“One facet of quality in research is that it meets the needs of the moment, or, ... is ‘timely’. Such research is more likely to be used and cited” (Rudd 1988:49). A determining factor therefore of the impact of research is the perceived research/knowledge needs at the time of assessment. A commonly used method of determining this type of impact is through the use of citation indexes, publication indexes, and peer review. For example, the indices used by the Australian National University (ANU) in their *Review of Quality* to determine the impact of research were “citations, patents, licences and major recognition through international/national invited lectures” (ANU 2003:3). “The ANU believes it is essential to look beyond measures of income and publications if a true indication of quality of research output is to be obtained. Accordingly, with the assistance of the ANU's Research Evaluation and Policy Project, research performance in terms of citations per paper and citations per staff member are measured” (ANU 2003:13). A common methodology for assessing the impact of research is through peer review. It is most important however, that peer review is complemented by performance indicators.

The Allen Consulting Group (2005:vii) note that to “...fully capture the impacts [aside from the academic and economic impacts already mentioned] of public research it is necessary to ask ‘what does society value’ and ‘does publicly funded research contribute to these dimensions of what society values?’. They (Allen Consulting Group 2005:vii) further provide four broad groupings of societal values

- *material – encompassing the goods and services available to society;*
- *human – encompassing physical health, mental health, quality of inner life, and the extent to which people can have pleasurable experiences;*
- *environmental – encompassing biodiversity and the quality of air, land, inlandwaters and seas; and*
- *social – encompassing social attachments, freedom from crime, the level and security of political rights and the extent to which the population engages in political processes.*

In relation to this, the Council for the Humanities, Arts and Social Sciences is conducting (Gilbert 2005:26)

...a study [funded by Department of Education, Science and Training] into ways of measuring the value and quality of research. [They note that it is] ...more difficult to measure quality in the humanities and social sciences in the absence of clearly defined output measures as in science. ... To evaluate research requires an analysis of three aspects. First, there is the methodology. This involves thoroughness and appropriateness. Academics are generally good at assessing these. Secondly, there is the assumed value of the research (unknown at the beginning). Thirdly, there is the impact (value) of the research results when recommendations are implemented. This is an area that tends to be neglected when the emphasis is on outputs. Let it be said, however, that not all research needs to have an immediate effect. It can be important to analyse simply to gain understanding even though you may not be sure where the analysis will lead. ... For

example, with the science department group, the aim (problem to be solved) was to determine how to bring together research institutions and industry. Counting the number of reports is therefore meaningless. You would assess the value of such research by the extent to which the ideas were, for example, useful to and taken up by industry.

2.2 Sustainability of Research

Sustainability of research, in the context of this paper, is taken to mean the ability for that research activity to be sustained in the medium-term. Factors affecting the sustainability of research might include such obvious factors as financial, infrastructure and research support. However, other factors that are not as readily identifiable might include the perceived need by funding agencies for specific research or research training in a particular field (a measure of this might be the number of post graduate degrees awarded in that field). Sustainability of research would seem to be related to the impact of research in that the ability for researchers to gain ongoing funds and investment will presumably be strongly linked to the impact of their past research. Also overall research program management will be a determining factor of the sustainability of the research, because without adequate leadership, management and marketing of the research, a research program that under other circumstances would be sustainable becomes unsustainable due to poor management. Also, in order for the research to be sustainable it needs to relate to institutional priorities. For example, an outstanding scientist might be able to sustain a particular line of research solely due to his/her standing in the scientific community, but if that person leaves the institution, the research will collapse unless it is built into the fabric of the institution. Moreover, issues of sustainability must be tempered with the realisation that science is dynamic and constantly changing. Current research programs need to be sustainable, but not to the degree that new, innovative areas of research are ignored.

2.3 Importance of Research

According to the Allen Consulting Group (2005: vi):

While there is merit in assessing the quality of research being conducted with public funding, and it may be appropriate for some types of research activity to allocate resources purely on the basis of research excellence, saying that high quality research is being conducted with public funds does not answer the fundamental and increasingly insistent question of policy makers of ‘what value is coming out of this research funding?’. To answer this question a focus is needed on the societal benefits of research, not simply the academic quality of research.

The value or importance of research will be linked to elements such as the relevance of the research to problems that exist at the time of the research, and possibly issues that are expected to occur in the future. For example, Universities are now expected to provide research that is nationally significant. Presumably indicators of this importance of research would include the extent to which the research meets an identified need. So even though a piece of research may not meet an identified need at present, it might be critical in the longer term.

The importance of research is linked to the sustainability of research, in that if the research is deemed to be important then presumably its chances of being sustainable would be greatly increased than if it was deemed to be unimportant. A problem exists here however, if the research that is deemed to be unimportant is in fact critical to our understanding of an unforeseen problem in the longer term. As such the perceptions of

research importance need to be understood and investigated to make an accurate assessment of research importance.

2.4 Potential of Research

Sornettee and Zajdenweber (1999:1) ask “at what level should government or companies support research?” Their answer is worth quoting at some length.

This complex multi-faceted question encompasses such qualitative bonus as satisfying natural human curiosity, the quest for knowledge and the impact on education and culture, but one of its most scrutinized component reduces to the assessment of economic performance and wealth creation derived from research. Many studies report evidences of positive economic benefits derived from basic research ... In certain areas such as biotechnology, semi-conductor physics, optical communications ..., the impact of basic research is direct while, in other disciplines, the path from discovery to applications is full of surprises. As a consequence, there are persistent uncertainties in the quantification of the exact economic returns of public expenditure on basic research. This gives little help to policy makers trying to determine what should be the level of funding. Here, we suggest that these uncertainties have a fundamental origin to be found in the interplay between the intrinsic "fat tail" power law nature of the distribution of economic returns, characterized by a mathematically diverging variance, and the stochastic character of discovery rates. In the regime where the cumulative economic wealth derived from research is expected to exhibit a long-term positive trend, we show that strong fluctuations blur out significantly the short-time scales: a few major unpredictable innovations may provide a finite fraction of the total creation of wealth. In such a scenario, any attempt to assess the economic impact of research over a finite time horizon encompassing only a small number of major discoveries is bound to be highly unreliable.

At best, judgement of the potential of research is an ‘inexact science’. Judgements of potential must balance the short-term and the long-term, the mundane and the cataclysmic.

2.5 International Comparisons

An assessment of the research capacity in universities by the GUNI-AP group will by definition involve international comparisons due to its members being located in China, Australia, India, Indonesia, Japan, Korea, Malaysia and Thailand. It is important therefore that the strengths and weaknesses of such an approach are understood and methods of dealing with particular issues, such as differences in university structure, are explored.

The most common drivers for undertaking such a comparison include “improving the quality/focus of research and enhancing accountability for public funding” (DEST 2005a:45).

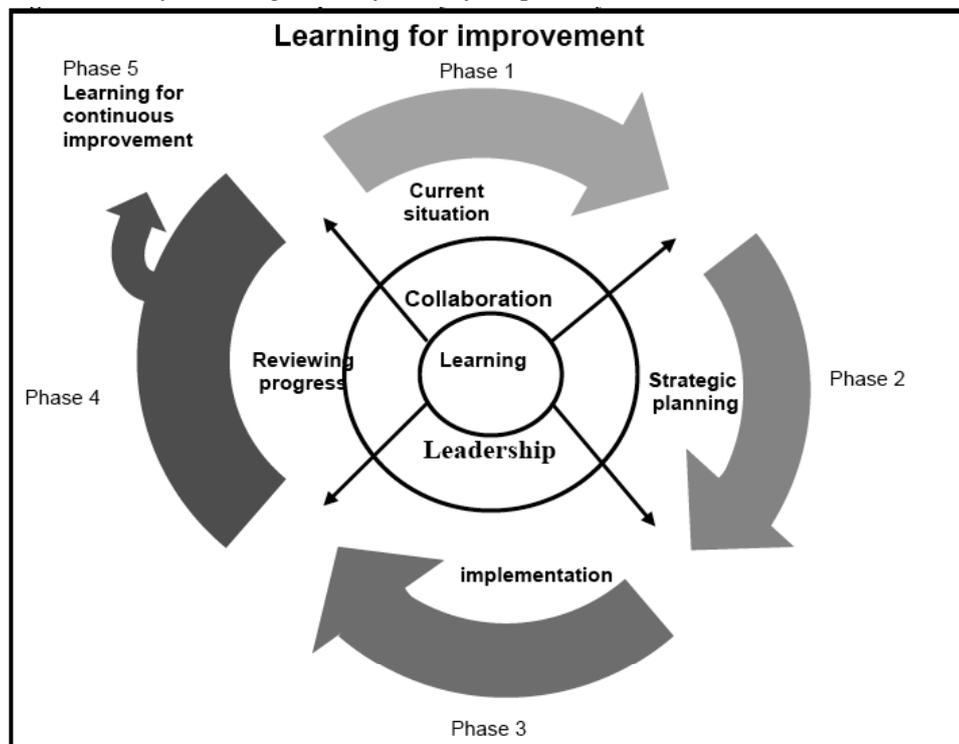
One of the main problems with benchmarking between universities however, has been a lack of consistency in the benchmarks used and the method of measurement, rendering comparison very difficult. *Benchmarking: A Manual for Australian Universities* prepared by McKinnon, Walker and Davis (2000) provided guidance here, however, this “one size fits all approach” has been criticised by Garlick and Pryor (2004:11) in that it does not recognize “...diversity in life cycle, location, governance, size and other variables in and between universities”.

Five phases were articulated as making up one generic cycle of working towards better practice. A number of sub-elements to each phase might be suggested; however, this would vary for each functional circumstance.

Figure 1 shows these five phases as:

- *comprehensively reviewing the current situation and environment as it relates to the targeted function*
- *undertaking a process of strategic planning targeted at improvement*
- *a program of implementation with the resource commitment of senior management*
- *a process of review to establish the degree to which improvement has occurred*
- *a recognition that learning from the previous phases can lead to further improved approaches in an ongoing process.*

The underlying principles of collaboration (or connectivity), leadership and learning are seen as influencing each of these five phases.



SOURCE: (Garlick & Pryor 2004:46)

Figure 1: Phases for an improvement cycle in university functions

2.5.1 Difficulties of International Comparison

The Allen Consulting Group (2004:21) found in their work comparing research quality in New Zealand tertiary education organisations to their British counterparts that:

A key limiting constraint on the conduct of international performance comparisons is the range of data that are available in relation to the research performance of universities across different international jurisdictions. If all TEOs (and all departments within TEOs) in all countries collected and publicly released research performance data using common data categories and collection methodologies, conducting valid and reliable comparisons of performance would be a relatively

straightforward exercise. In practice, however, performance data collected and publicly released tends to differ considerably across (and even within) jurisdictions. Therefore, prior to consideration of how data can be used to make international comparisons, it is important that what data are available is understood. Available data can be loosely grouped into seven categories, namely:

- *bibliometric data;*
- *awards to individual researchers;*
- *research student data;*
- *research faculty data;*
- *research income from external sources;*
- *research commercialisation performance data; and*
- *outcomes from peer review processes.*

It is also important to note at the outset that some measures of research quality may not be universally applicable across research fields. For instance, even the most commonly used existing bibliometric measures of quality such as citations have varied levels of acceptance across different fields of research. While broadly accepted as a valid quality indicator in the physical and natural sciences they are not as widely accepted as valid indicators of quality in the humanities and social sciences community.

They (Allen Consulting Group 2004:21) further caution that

Given that the introduction of international comparisons that are not valid and reliable could carry significant negative consequences for the tertiary education sector, the introduction by the TEC of a system for international performance comparisons must be handled sensitively and should only proceed once the following three steps have been completed. However, while the steps required for the TEC to introduce a valid and reliable system of international comparisons of research performance would be the same in principle, in terms of their application it would appear that they would take longer to apply and require considerably greater sector consultation the higher the complexity and stakes involved.

The three steps required would appear to be that:

- *firstly, it needs to be determined, and broadly agreed between stakeholders, what the core purpose(s) for conducting international comparisons of research performance should be. The purposes for the comparisons will play a significant role in informing the selection of an appropriate approach to making international performance comparisons;*
- *secondly, wide consultation should occur on what approach or approaches to performance comparisons will best allow for the purposes of the exercise being met; and*
- *finally, in consultation with stakeholders, a detailed methodology and implementation plan will need to be developed that allows valid and reliable international performance comparisons to be efficiently made in a way that allows the objectives of the exercise to be met. The complexity in practice of this final step would appear to be significantly dependent on the outcomes of the first two steps.*

2.5.2 Examples of international benchmarking

Williams and Van Dyke (2004), of the Melbourne Institute of Applied Economic and Social Research, recently carried out an exercise which compared and ranked Australian Universities against their international counterparts. They used a combination of both

quantitative and qualitative measures, with the aim of providing comparative information to prospective students, academics and employers.

Other recent examples of international comparison include that of the *Times Higher Education Supplement* in the United Kingdom which compared universities in terms of overall standing. “The ranking, employing a 50-50 split between qualitative peer rating and quantitative measures, compares institutions on a weighted assessment of reputation, research and teaching dimensions”. Alternatively, “the Institute of Higher Education of Shanghai Jiao Tong University attempts an international ranking of universities solely using bibliometric measures to compare universities on academic and research performance” (DEST 2005a:43-44).

As would be expected, each of these exercises used a different set of criteria upon which to base their assessment. A summary of these is provided in Table 1.

Table 1: Criteria used in each of the international assessment exercises

Criterion	Melbourne Institute	THES	Shanghai Jiao Tong
Quantitative peer review, eg. awards, citations	40%	20%	90%
Qualitative peer rating, eg. surveys	8%	50%	
International staff		5%	
International students		5%	
Graduate programme quality	16%		
Quality of undergraduate entry	11%		
Undergraduate programme quality	14% ^a		
Staff/student ratio	a	20%	
Resource levels	11%		
Size of institution	b	c	10%
<i>Total</i>	<i>100%</i>	<i>100%</i>	<i>100%</i>

a Melbourne Institute of Applied Economic and Social Research uses staff/student ratio as one of four measures in assessing the quality of undergraduate programmes

b Research activity is measured both in levels and with a size correction then combined as a weighted average

c International students, international staff, citations and staff/student ratio are all measured as ratios

SOURCE: (DEST 2005a:44)

DEST (2005a:45) goes on to comment that “National rankings may be important in some contexts, but to know how well Australia is doing in a particular area requires international benchmarking”.

2.6 Criteria of Quality in Research

“Currently there is no robust and consistent way to measure the quality of research conducted in universities and publicly funded research agencies and its benefits to research and the wider community” (Australian Government 2004:online). However the Australian Government has very recently released an issues paper, *Research Quality Framework: Assessing the Quality and Impact of Research in Australia* (see DEST 2005a), which seeks to provide guidance on this issue. This is just one example of a solution developed within a particular country or institutional setting that adds yet another framework to the mix, but does not provide a universal solution. In fact, it is futile to

even contemplate that such a solution might exist. In this increasingly complex world, context provides us with the framework within which solutions should be developed. The best we can do then is to assimilate the numerous frameworks and tools that are available and apply them to the situation in which we find ourselves in a way that is going to inform our operations

The use of performance indicators to benchmark University research performance in many countries has now become standard practice, particularly given that these measures have become increasingly linked to the allocation of public funds (see Carotenuto et al. 2001; Johnson & Osborne 1997; McNay 1998; Ramsden 1999; Tomlin 1988). A survey of the literature indicates that current practice relies heavily upon 3 main groups of indicators, that is research income, publication output and higher degree completions (see DEST 2000; Joint Working Party on Higher Education Indicators 1994; Ramsden 1999).

There are essentially two categories of performance indicator, these are quantitative (such as the number of publications) and qualitative (such as the importance of the publication). Both types have their advantages and disadvantages, however the message here is that they should be used to complement each other, they should be applied thoughtfully and within the particular context of the organisation and in a way that enables their measurement to feed into a continuous improvement cycle.

2.6.1 Measurements of Quality

Rudd (1988) presents a useful discussion in this area, particularly in relation to the context in which it is appropriate to evaluate quality in research, the criteria of quality and the types of measures that might be used. He (1988:48) describes the main criteria used in the evaluation of quality in research as:

- a. subjective evaluation (peer-review);
- b. numbers of publications;
- c. citations;
- d. the number and value of research grants and contracts gained.

A brief description of each of these will be presented in the following.

2.6.1.6 *Subjective evaluation: peer-review*

Peer review is the most widely method used to judge the quality of research (Rudd 1988) in a variety of settings including research funding applications, articles submitted for publication and job applicant selection. “The criteria used may include, in addition to the likelihood of the proposed research making a significant and original contribution to knowledge, the extent to which the research fits into an overall programme of research, or is concerned with certain priority areas, or, for example, in the Netherlands, shows ‘social merit’” (Rudd 1988:49).

An advantage of the approach is that “a well informed insider will be able to spot trends far sooner than the impartial outsider. So one has to conclude that peer assessment has a value, and that due weight must be given to the opinions of people who actually understand the subject being examined” (Rudd 1988:49). There are also a number of difficulties with the peer-review method of assessment in that “a simple mathematical sum of the grades awarded by referees may produce the wrong answer. Whatever the instructions, the referees may not follow them, especially if they disagree with them” (Rudd 1988:49). “The key problem is how to find committees who are truly detached and disinterested, and referees who combine knowledge of the area in which applicants hope to work with sound judgement and impartiality” (Rudd 1988:49). For example, the

referee may view an applicant as a potential competitor for funds, or that the applicant is using an inferior methodology to that which the referee might use.

There is some debate in the literature over whether performance indicators or peer review is the most valid method for assessing research performance in particular disciplines (for an extended critique of peer review see Chubin & Hackett 1990). The advocates of peer review maintain that it is the disciplinary expert who is best placed to make judgements about quality in his/her area of research expertise. Peer review is a fundamental aspect of the academic process and it is the internal professionals who must ultimately judge and be held responsible for the quality of the knowledge they produce and manage. But the use of research performance indicators/bibliometrics (see following section) and peer review is not mutually exclusive. Performance indicators and bibliometrics actually incorporate the peer review process in that before an article is published in a refereed journal, its contribution to knowledge is assessed by 'peers'. Also, where one of the indicators measures success in obtaining external research funding, peer review is brought into the formula in that research agencies often allocate grants through an assessment procedure involving peer judgements. A more sophisticated approach maintains that performance indicators and peer review should complement one another and, hopefully, help counterbalance the inherent biases in each individual approach (Sizer 1990a). Furthermore, Sizer (1990b:155) argues that "peer review should not be contrasted with performance indicators"; they are complementary approaches, not polar opposites. Moreover, as will be discussed in more detail below, in many areas, such as research, it is impossible to divorce numerical measures of performance from more subjective peer judgements.

2.6.1.7 Number of Publications

As nations increasingly are having to formulate appropriate and effective research policies, they are turning to the development of reliable techniques for monitoring research outputs, both in relation to the production of knowledge and the use of research for economic development (Irvine & Martin 1985).

Productivity in research, however measured, is bound to be influenced by the extent to which the individual has succeeded in gaining funds for research, and, in as far as the allocation of funds is influenced by past productivity, ..., i.e. those who have already achieved recognition are more likely to have their work read and to be credited with further significant contributions to knowledge. It follows that those researchers who have had funds in the past can get them more easily in the future ... anyone without funds is unlikely to achieve substantial numbers of publications. So what is being measured when publications are counted is partly the ability to gain funds, which is generally in itself partly the result of a judgement on the quality of the individual's research (Rudd 1988:50-51).

Rudd (1988:51) further notes that "there is an assumption that everyone has an equal chance of getting his [or her] results into print", however, there is evidence to suggest that this is not always the case due, for example, to factors relating to the authors notoriety and institution.

Due to this focus on the number of publications and their type i.e. whether they are refereed or not has led to an academic culture of 'salami publishing', where one piece of work is divided into several papers, or by submitting essentially the same article to different journals (see Rudd 1988).

Many of the techniques used to measure the number of publications are based on the statistical analysis of publication rates and citations (see next section), or what is known as bibliometrics. Several commentators have noted that it is necessary not to confuse the technical requirements of the bibliographic approach with the more general concept of performance indicators (Johnes 1988:56; Rudd 1988). Nonetheless, bibliometrics is becoming a popular mechanism for assessing research performance and thus the method and its limitations deserve consideration.

Alan Pritchard coined the term 'bibliometrics' in 1969 and defined it as “the application of mathematics and statistical methods to books and other media of communication” (ASTEC 1989:329). A more recent definition is “the quantitative measurement of the properties of a literature, usually as an aid in exploring the production, distribution and utilisation of its contents” (ASTEC 1989:329). A bibliometric statistic becomes a performance indicator when it is infused with value or worth.

While performance indicators produce a set of empirical data that can help inform management and policy decisions, those working in the area often caution that no one indicator, such as publication rate, paints a true picture of the research performance of a particular department or unit, and that the data should be approached with caution and supplemented with more qualitative type information (Tognolini, Adams, & Hattie 1994). They also caution that research performance indicators should not by themselves be used to determine funding levels. But performance indicators have great appeal to the hard-pressed bureaucrat for “they readily lend themselves to comparison and thus can be used to determine and to justify resource allocation decisions” (Brennan 1990:110).

Some difficulties with using citations as a measure of research impact exist, for example Rudd (1988:52) states that “in theory, counts of citations ought to be a more effective measure of productivity in research than counts of articles, even where these make some attempt to distinguish between the more selective and the less selective journals”. However, “problems arise (a) from the assumption that citing a book or article shows it is a useful contribution to knowledge and that the most useful works are cited most frequently; (b) from the extent to which, in citing other works, authors actually behave as they are expected to behave; and (c) from the limitations of the published counts of citations” (Rudd 1988:52). Indeed “the frequency with which a work is cited is far from an accurate measure of its importance” (Rudd 1988:53). Finally, “the use of citations as an indicator of the quality of research is heavily biased in favour of pure research and against applied research” (Rudd 1988:53) because applied researchers are less likely to publish.

Furthermore, citation indexes also report on the ‘impact factor’ of a particular journal, this provides a measurement of the prestige and international visibility of journals” (Bordons, Fernandez, & Gomez 2002:195). This is of particular importance where the national journals of perceived marginal countries do not appear in the indexes such as ISI. As a result this has strongly influenced the publication strategy of academics toward those journals that are deemed to have a high impact.

2.6.1.8 *Research publications index*

The research publications index is designed to “recognises quality "low-cost" research, particularly in the Social Sciences and Humanities” (Bond University 2003:2-3). The Research Publications Index is a complex formula comprised of the following variables: Books, Chapters, Refereed Journals, Refereed Conference Articles, Creative Works, Patents, and Refereed Designs.

2.6.1.9 *The number and value of research grants and contracts gained*

The number and value of research grants and contracts gained is thought to be a good proxy for research quality, presumably because one must have good research capacity to win a research grant or contract. For example, Tognolini, Adams and Hattie (1994:112) state that the value of research grants “is a reflection of the quality of the department, as granting bodies give money only to established high-quality researchers who typically have excellent track records in producing important research”. As with other measures discussed thus far, “it is past success in research that determines current success” (Rudd 1988:54). It is apparent though that the results for such a measure would be skewed toward those researchers/research groups that have had previous success, where as up and coming researchers/research groups will under perform in this category because they are as yet not established and thus will in all likelihood not have a great number of research grants and contracts. Furthermore, “in some departments, a low level of research income does not necessarily mean a low level (or quality) of research” (Tognolini, Adams, & Hattie 1994:113). The use of the value of research grants as a measure of research quality is therefore not as straightforward as one might initially think. It must also be used cautiously and one must consider all perturbations that the data might mask.

It is thought that the number of research grants as opposed to the value of the research grants might be a ‘fairer’ measure of research success, as it allows variations between subject areas. However, the data would need to be normalised according to average research grant value within the various disciplines (see Tognolini, Adams, & Hattie 1994:113 for further details). This could lead to the proliferation of small grants.

2.6.1.10 *Quantitative vs Qualitative Indices*

The other thing to consider is whether the indicators should be quantitative, as is currently the case in Australia, or qualitative in nature such as those used by the Research Assessment Exercise (RAE) in the UK.⁵ Table 2 describes some advantages and disadvantages of the quantitative performance indicator model and the qualitative peer review model.

Charles and Benneworth (2002:4) outline the disadvantages of both quantitative and qualitative indicators. First, with respect to the disadvantages of quantitative indicators they maintain that:

- a. *They reflect past actions and policies rather than current strategies.*
- b. *They are highly influenced by the structure of HEIs and inputs such as the quality of students.*
- c. *They are often crude surrogates for what needs to be measured, and there is a risk of seeking to deliver the required indicators rather than the desired outcomes.*
- d. *It is easier to measure absolute outputs than value added.*
- e. *There may be significant time-lags in ultimate success.*
- f. *Significant economic impacts may require risk-taking, and hence there is a likelihood of short-term failure and poor performance even if the approach is correct.*

Second, the shortcomings of qualitative indicators include:

⁵ The RAE exercise allocates a greater proportion of research funds, 60% of all public research funds flowing to Universities, as opposed to 11% through the quantitative measures employed by Australian Universities (Ramsden 1999: 342).

- a. *Good practices may depend on the context, so the relative success of one approach may be difficult to judge.*
- b. *Most qualitative evaluation is akin to description, and generalisation is difficult.*

Performance Indicator Model	
Advantages	Disadvantages
<ul style="list-style-type: none"> • Incentives for research productivity • Incentives to recruit and retain high performing staff • Incentives to secure external research funding • Incentives for quality research degree supervision • Data easy to collect • Formulaic and therefore transparent • Low transaction costs 	<ul style="list-style-type: none"> • Indicators are only a rough proxy for research excellence, and some focus too much on quantity rather than quality • May create goal-displacing behaviour • Some disciplinary areas or institutions (i.e. those who traditionally get more of the limited external research funding) will be favoured at the expense of others • Only modest incentives for cooperation or collaboration
Peer Review Model	
Advantages	Disadvantages
<ul style="list-style-type: none"> • More direct and accurate measurement of quality • Enables centres of excellence to be identified • Assessment methodology allows factors other than peer review to be considered • Allows inclusion of international perspective (on panels) • Effective means for concentrating resources in areas of strength • May be used for international benchmarking • Scalable, multifunctional and flexible 	<ul style="list-style-type: none"> • Higher administrative and compliance costs • Little weight to the views of end-users • May incentivise perverse staff hiring policies – i.e. competition for 'stars' • May discriminate against women academics • May entrench 'academic traditionalism' and penalise non-mainstream research • May be difficult finding adequate peer reviewers in small New Zealand system

Table 2: The advantages and disadvantages of two evaluation models

SOURCE: OECD (2004:13-14 in DEST 2005a:50-51)

The authors conclude that:

Both quantitative and qualitative assessments suffer from a problem of scale. At what scale should measurement take place, and what measurements can be applied across a wide range of departments and activities? Many activities are undertaken by parts of the HEI only, or are small elements within the work of a department, and hence tend to be unmeasured or unreported to the centre of the HEI. The efforts of individuals may be significant but go unnoticed within their departments. (Charles & Benneworth 2002:5)

One would logically envisage that the rank order of institutions might vary depending upon the criteria used for evaluation: publications, citations, number or value of grants, quantitative or qualitative indices, etc. However Irvine and Martin (in Rudd 1988) found in their research that that whichever criteria were used the outcome was reasonably consistent. In part, this may be due to the fact that “there is a tendency here (as elsewhere) for only the variables that can be measured, and so can be put on the computer, to be taken into account” (Rudd 1988:54).

Clearly, there are problems associated with the use of all performance indicators. Recognition of that fact, however, is not an argument against their use, but rather a note of caution as to how, where and why they are employed.

2.7 Examples of Research Quality Assessment Processes and the Indicators Used

A summary of some of the major changes taking place in several countries regarding research assessment is provided by Von Tunzelmann and Mbula (in DEST 2005a:43).

Major findings were that many countries are increasingly relying on self-evaluation (subject to oversight panels). There are also problems (eg cost) with applying models which have been developed for large nations (such as the United Kingdom) to smaller countries. Other critics of research assessment have argued that the processes have been too focused on academic 'excellence' and excluded many aspects of research outcomes, particularly those related to industry and social impact.

There are a range of models of research assessment which have been developed overseas. The United Kingdom Research Assessment Exercise (UK-RAE) and the New Zealand Performance Based Research Fund (NZ-PBRF) are probably the most widely known approaches in Australia but there are also different models which have been developed or are being developed in The Netherlands, Germany and Ireland.

2.7.1 Higher Education Research Data Collection - Australia

The Australian Vice-Chancellors Committee (2003:19) notes that:

Higher Education Research Data Collection (HERDC) provides a useful mechanism by which universities can assess their own research performance and benchmark against other institutions. The categories involved – research income and research publications – provide, along with data on research student numbers and completions, a proxy measure for overall university research performance and output.

The broad groupings of indices they use in their collection are (AVCC 2005):

- University Research Income;
- University Research Publications; and
- Research Training: Masters by Research and PhD completions.

Further information regarding the detail of these broad groupings are contained in Appendix Two (A2.1).

These indicators are used primarily to inform institutional funding allocations through the Institutional Grants Scheme (IGS) and the Research Training Scheme (RTS).

Research funding agencies, policy makers, Government and universities all use the HERDC for a variety of purposes, despite mutual recognition of its forgivable shortcomings. Overall, the HERDC is suitable for the purpose to which it is put, although universities have some concerns about changing data requirements for information relating to commercialisation of research outcomes. (AVCC 2003:19)

2.7.2 Australian National University Research Quality Review 2004

Recently, the Australian National University completed

a review of the quality its research; research training, undergraduate and postgraduate education, the effect of its regional and national service and the strength of its international engagement. Part of the review process involved a research quality audit based on the UK-RAE. By using peer review and peer perception, the ANU believes it has obtained an informed external review of its performance that can be used as a guide to develop strategies for its pursuit of international excellence. (ANU 2004 in DEST 2005a:54-55).

2.7.3 The Research Assessment Exercise – United Kingdom

The Research Assessment Exercise provides qualitative measures about the quality of the research, based primarily on “...how much of the work is judged to reach national or international levels of excellence... [T]he research submitted is assessed against a benchmark of international excellence in the subject concerned” (Research Assessment Exercise 2001:2).

One of the early criticisms of the REA was that “since publications were seen as the main basis of panel judgments ..., staff were encouraged to publish for their peers in academic journals rather than for practitioners in professional journals...” By 1996, the REA had “countered this by not requiring total publications to be listed but only the four best outputs per person” (McNay 1997:37).

The RAE has been criticised as inhibiting institutional entrepreneurialism. The “entrepreneurial approach links research to industrial liaison, consultancy, and enterprise initiatives and there is a search for distinctiveness, for market niches, often through interdisciplinarity – ... something which [the] RAE was seen as discouraging, given its disciplinary base for quality assessment” (McNay 1997:38). Other advantages and disadvantages of the RAE include:

On the positive side, it is argued that the RAE has provided the basis for increased research funding and more comprehensive information on the quality of research being undertaken by subject area.

However, on the negative side, there are concerns regarding the financial sustainability of research in some disciplines, inequitable workplace behaviours (‘game-playing’ related to determining ‘active’ research staff), administrative burden and costs, and a need to fully recognise all aspects of excellence in research, for example pure intellectual quality, value added to professional practice, applicability and impact within and beyond the research community.

There has been some commentary on the possibility of introducing a UK-RAE or variant to Australia. For example, in Bourke’s examination of both the Australian and UK experience of evaluating research outcomes for the purpose of performance-based funding allocation, he recommended a scaled-down Australian model of RAE using national discipline reviews operating at a university-wide field of research level. Similarly, the idea of a funding scheme based on a process in which the quality of a university’s research performance is assessed is seen as useful for the higher education sector as a whole. (DEST 2005a:47-48)

2.7.4 League Table Approach

The league table approach is merely ranking institutions on one or more indicators according to prestige. One of the main conclusions about this league table approach reached by Ball and Wilkinson (1994:426) was that the “league table mentality” was inappropriate in that “[h]igher education institutions are not football teams, they have

differing institutional missions and operate in different environments”. They do however offer an alternative approach to the league table approach, which involves the presentation of institutional policies and associated performance indicators which would then enable an assessment of institutional performance.

PART 3: Designing a Research Evaluation Process

It is not the intention of this paper to provide a prescriptive set of processes and indicators for casual application by institutions. Rather we have provided a set of issues to be mindful of in designing a research evaluation process, summaries of performance indicator types and most importantly we have outlined a number of processes that might be employed in undertaking a research performance assessment.

There are many factors to consider, many of which have already been mentioned in this paper, when devising a set of indicators to ‘measure’ research performance within and amongst Universities. This might include the definition of the types of universities or the disciplinary cluster for which the particular suite of measures are valid. Consider firstly as an example, the recent legal battle that preceded the impending release of a university league table for New Zealand (which compared a number of elite British universities with a local group that were considered to be their counterparts) as the international comparison was thought to be “analytically flawed” because the comparison was not valid (Cohen 2004:para 2).

Secondly, Professor Malcolm Gillies, President of the Council for the Humanities, Arts and Social Sciences (CHASS)⁶ stated in a recent media release that “[D]ifferent researchers produce different products. Some write books and papers, others contribute ideas at conferences. ... There is also the largely unexplored world of researchers working in music, art and performance, and the equally unexplored area of contributions to public policy. The aim of the project is to develop a new set of measures which will pick up a variety of different research outputs, and help funding agencies identify the best” (CHASS 2005:online). The suite of indicators should thus “be sensitive to different quality measures for different disciplinary clusters” (Illing 2005:31).

The suite of indicators will therefore, by necessity, fall somewhere along a continuum between ‘lowest common denominator’ coarse grade indicators and a fine grade suite of indicators that are able to detect and allow for the slightest variation. The former end of this spectrum is of limited utility and the latter probably not achievable in practice. Most efforts at benchmarking research performance using performance indicators will fall somewhere in between these two extremes. The basic point is that:

Evaluation criteria need to recognise the quality of the research, its potential social and economic impact, and the value of university research in educating students. In this area, national initiatives should be complemented by further efforts at international level to develop benchmarking indicators and methodologies, and promote the use of foreign expertise in national evaluation. (OECD 2002: 10 in DEST 2005a:43)

A number of elements that need to be considered in the design of an evaluation process are outlined below. These are by no means an exhaustive list, however they do represent the core elements that need to be taken into account.

⁶ CHASS is currently undertaking a project to design measures that are valid indicators of research performance in the humanities, social sciences and arts.

3.1 Purpose of Review

Tognolini, Adams and Hattie (1994) developed a methodology to aid in the selection of performance indicators for research quality. They advise (1994:105) that “the procedure for identification and the definition of performance indicators need to be reconciled prior to the implementation of models of performance indicators, if the model is to have a high probability of becoming accepted within a university context”. This would require as the first order of business a process through which agreement is reached within the university regarding “...the aim of the exercise, the research aims of the university, and what constitutes research in the university” (Tognolini, Adams, & Hattie 1994:115). They identified a total of 82 indicators with which to measure research performance from the literature which they then asked departmental heads to rate the indicator according to “...its relative importance for measuring the research attainment of their departments” (Tognolini, Adams, & Hattie 1994:106). The final suite of indicators were thus chosen according to the preference of the departmental heads therefore maximizing agreement and should therefore reflect the diverse nature of research across disciplines.

3.2 Alignment with Mission

The primary recommendation by numerous authors (see Andrews et al. 1998; Ball & Wilkinson 1994; Dochy & Segers 1990) in relation to performance indicator development and selection is that they be consistent with the mission of the organisation in question. However, this begs the question, how valid is a comparison between these organisations? “Since it is valid, indeed highly desirable, for different institutions to have different objectives, comparison of institutional performance by performance indicators would be misleading” (Ball & Wilkinson 1994:418-419). They (1994:426) further argue that:

Diversity of institutional missions is clearly beneficial and thus provides a range of institutions that will give a genuine choice to prospective students in particular and society in general. Thus it is essential that an institution gives considerable thought to its mission and objectives, how to measure performance and which performance indicators are most appropriate for this purpose. This may of course involve discussion and agreement with the appropriate funding body. The advantage of the approach is that each institution should “own” its performance management process.

It follows then that institutions should make it clear what indicators should be targeted solely at internal review processes and what indicators are to be used for external comparisons.

3.3 Regional Relevance

Universities are no longer simply educational institutions, they now have a responsibility to, amongst other things, provide research that is nationally and regionally relevant. It is therefore also important that benchmarks and performance indicators are developed to enable universities to determine their contribution to economic and social development of their region.

An excellent example of a research institution that annually undertakes an evaluation of their research performance, particularly in relation to their influence on their region is Manaaki Whenua — Landcare Research located in New Zealand. For example, much of their Annual Report for 2004 (see Manaaki Whenua - Landcare Research 2004) is dedicated to illustrating the relevance and impact of their research in their region, to their nation and internationally. Much of this is achieved through storytelling, rather than

simply providing measures of success. They also provide performance indicators to quantify their impact, however this is not the focus of the report.

3.3.1 How can it be measured?

Whereas McKinnon et al. virtually ignored the external relations between the university and its regional community, Charles and Benneworth (2001) identified seven broad categories and 34 subcategories of regional development outcomes from a university's partnership performance with its regional community. The categories were regional governance, human capital development, economic competitiveness social development, sustainable development, cultural development and equity issues. (Garlick & Pryor 2004:18)

Charles and Benneworth (2002) developed a benchmarking tool to enable higher education institutions to assess their regional contribution with reference to the needs of the regions in which they operate. The intention being that institutions will be able to “identify priorities and work towards good practice” (Charles & Benneworth 2002:3), rather than as a mechanism to allocate funds within the institution through ranking.

The benchmarking tool they developed has three functions (Charles & Benneworth 2002:2):

- to assess improvements in the strategy, performance and outcomes of [higher education institution] HEI regional engagement
- to help the HEI set its strategic priorities
- to support joint strategies within a regional partnership.

The objective is to give individual HEIs a means of assessing their regional impact. The key challenge is to highlight not just linear relations between an HEI and its region, but also a wide range of strategic interactions. Strategic priorities for regional engagement should be regional development processes which link between, for example, economic development and educational attainment, or community regeneration and the formation of new firms. (Charles & Benneworth 2002:3)

Their benchmarking tool consists of 33 benchmarks (these have been reproduced here as Appendix Three) which can be categorised as either focused on practice or performance. This distinction then enables an examination of the spread of engagement across the range of regional development activities “and where the balance is between current performance and those practices needed to ensure improvement in future” (Charles & Benneworth 2002:6).

3.4 Scale

There are important decisions that need to be made early in the process of indicator selection regarding the scale upon which data is required eg. inter-university, departmental or individual. This will then guide the types of indicators that are appropriate for use at each of the scales that an institution is interested in.

3.5 Indicator selection

This section outlines a number of factors that need to be considered in the selection of individual indicators and groups of indicators. This is by no means an exhaustive list, but does cover some of the main issues.

In their report *Characteristics and Performance of Higher Education Institutions*, Andrews et al. (1998:1) presented five criteria that they considered to be desirable in the selection of indicators. These were:

- Relevance – related to the purposes and objectives of the institution;
- Reliability – based on consistent data;
- Auditability – data are able to be scrutinised;
- Transparency – data are to have clear meaning; and
- Timeliness – data to support the indicators are currently available.

They further argued that “[t]he development of performance indicators needs to acknowledge the differing missions and circumstances of institutions. Nonetheless, higher education institutions share a core set of purposes related to teaching and research and need to satisfactorily perform against these purposes” (Andrews et al. 1998:1).

Further to this, we would argue here that the purpose, availability and cost of each performance indicator also needs to be considered during the selection process.

- Purpose – what will this PI measure and how will it indicate required change;
- Availability – how straightforward is it to collect the relevant data; and
- Cost – how much will it cost in terms of time and money to collect the data.

It has also been suggested by some authors (eg. Cave et al. 1998) that indicators should be chosen to evaluate the whole research process, this would therefore include indicators about research inputs, processes and outputs. In doing so then all phases of the research process are represented in the assessment, rather than simply outputs which seem to be the current focus. Some indicators might cut across these categories, however the point is not which categories they fit into but that each of these phases are represented in the evaluation process.

3.5.1 Inter-university comparison and data standardisation

Ball and Wilkinson (1994:427) argue that “[t]he safest and fairest conclusion is that performance indicators are an invaluable even essential tool of management information within institutions; they are misleading and unhelpful tools of political analysis between institutions – unfortunately the latter are more exciting”. It is apparent then that when undertaking benchmarking between universities that :

- representatives from all universities are involved in the indicator selection process and to identify which apply to their institution and those that do not rather than using a ‘formula’ or ‘cook book’ approach;
- there might be a tiered approach to indicator selection for example:
 - Tier 1: those that apply to all universities (the data from these will need to be standardised to enable fair comparison - see Ramsden 1999 in relation to data standardisation);
 - Tier 2: those that might apply to only a subset of the institutions within the group;
 - Tier 3: those that apply within individual universities only.
- there may also be a need for publication restrictions to be placed on access to particular data sets, for example some data may only be appropriate for internal review and not for wider publication.

3.6 Action Plan

This paper has not attempted to specify particular performance indicators, for it is assumed that it does not make sense to discuss a particular indicator outside the context in which it is to be developed and implemented. It is however within the scope of this paper to recommend a process that the GUNI-AP member universities might use as a starting point for their research quality and assessment process.

It is obvious from the arguments presented in this paper that the process might move through the following phases:

1. Commitment by management across the GUNI-AP universities to commit to the assessment exercise and the reflect upon the feedback it provides if the aim is to promote institutional strengthening and continuous improvement.
2. Agreement reached between the universities about the aim of the assessment exercise.
3. Agreement within the universities about the aim of the exercise, the research aims of the university, and what constitutes research in the university
4. Agreement on the types of indicators to be used across all the GUNI-AP universities, particularly in relation to those that will be used specifically between and within institutions.
5. Individual universities may develop their own set of indicators for use in internal review processes.
6. The GUNI-AP universities will have to decide how much of this information they would like to release into the public domain. Some could be kept as commercial-in-confidence, while other potentially less favourable data could be used for internal review. In general, however, we would argue for openness and transparency.
7. Set in train a review cycle (probably after the first review) that:
 - a. re-examines the aims of the exercise;
 - b. examines the effectiveness of the indicators suite chosen in providing information that meets the aims of the review;
 - c. determine whether the information provided in the review has improved understandings about the function of research in universities and reflect upon whether as a result of this 'new' understanding different indicators might need to be considered;
 - d. determine whether the context has changed across/between or within institutions. For example if the mission of a university has changed to focus more exclusively on research then that needs to be reflected in the types of indicators it uses to benchmark against.
8. Begin new research evaluation process.

PART 4: Conclusion

This paper has not attempted to specify particular performance indicators or benchmarking techniques, for it is assumed that it does not make sense to discuss a particular indicator or technique outside the context in which they are to be developed and implemented. Moreover, there is no objective method available that can say that one indicator is better than another in a context-free situation. The appendices list some of the common indicators used in various benchmarking studies. These lists should be used only as a basis of discussion and further consideration. A useful performance indicator will only be derived from negotiation amongst interested parties in a specific context. In a situation where objectives themselves are under dispute, which is often the case in higher education, negotiating agreement on a commonly acceptable set of performance measures may be a time consuming and difficult task.

Part three of the paper suggests an approach to the use of performance indicators in benchmarking research performance amongst the GUNI-AP universities. As indicated above, the first task is to determine what are the core purposes for conducting comparisons of research performance amongst the group. This probably must first be determined at the individual institutional level through consultation with all relevant stakeholders. The areas identified in need of performance assessment must at the very least relate to the institutions' basic mission and objectives.

Once individual institutions have determined their specific objectives for research performance comparisons, then the next task will be to identify which of these objectives the group hold in common and lend themselves to collective comparisons. These areas might then form the basis upon which to plan the more detailed approach to the actual benchmarking exercise.

The next task will be for the group to decide what comparative benchmarking approach best meets their common purposes. Here, it will be very important to take account of individual institutional contexts to ensure that like is being compared with like. For example, if one of the objectives is to compare scholarly output of academic staff in a particular disciplinary area, it would make little sense to measure the number of scholarly publications of academics in an older, well endowed research oriented university with that in a newer, primarily teaching only oriented university. If the group selects disciplines as the basic unit of measurement of, for example, number of publication or grant money attracted, then the number of academics within disciplines amongst the universities will need to be moderated using some agreed upon normalization process.

A good deal of effort will need to be devoted to identifying similarities and differences amongst this group of universities, particularly where it can be expected that such factors will have an impact on research performance. Some of the questions that might be considered include: what is the structure of research support in the respective countries; how do the various institutions structure research groupings (disciplinary based departments, interdisciplinary schools; independent research centres, etc); who controls the research agenda (faculty, deans, central managers, etc); how is the research budget decided; are all staff expected to engage in research as well as teaching; and so on. This exercise will also help to determine the basic units of measurement(s), which are probably likely to be institutional sub-units, such as disciplines, rather than whole of institution comparisons.

Finally, the detailed methodology for the benchmarking exercise must be decided. This will relate to the overall purpose of the exercise, but will also involve such practical issues

as the amount of time and resources that can be devoted to the exercise. The performance indicators selected will need to suit the chosen methodology. For example, disciplinary-based peer assessment of research performance would not suit a primarily quantitative oriented approach. Methodology will also depend on the cycle of review: annually, once every five years, etc. Where peer review is employed, at least some peers would need to be selected external to the GUNI-AP group.

Whatever the particular benchmarking methodology selected, it is likely that at least some of the common performance indicators will be employed in one form or another: number and value of grants, citations and other bibliometric indices, patents, awards, etc. But when using such measurements, as we stressed earlier in the paper, a clear distinction needs to be drawn between the quality of the research and the impact of the research being assessed.

The benchmarking of research performance amongst the GUNI-AP group of universities must be developed and owned by the group, individually and collectively. It is the users of the information that are important, not the information per se. The purpose of this paper has been to provide background information to assist the group to achieve this end.

REFERENCES

- Allen Consulting Group 2004, *International comparison of research quality in the New Zealand tertiary education system: Discussion Paper*. Retrieved 11 February 2005, from <http://www.allenconsult.com.au/nztec/>
- Allen Consulting Group 2005, *Measuring the Impact of Publicly Funded Research, A report to the Department of Education Science and Training*. Retrieved 09 May 2005 from http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/research_quality_framework/publications.htm
- Andrews, L., Aungles, P., Baker, S., & Sarris, A. 1998, *Characteristics and Performance of Higher Education Institutions (a preliminary investigation)*. Canberra: Research Branch, Higher Education Division, DETYA.
- Ashworth, A., & Harvey, R. 1994, *Assessing Quality in Further and Higher Education*, London: Jessica Kingsley.
- Australian Government. 2004, 'Quality and accessibility frameworks for publicly funded research', *Backing Australia's Ability: Building Our Future through Science and Innovation*. Retrieved 10 February 2005, from http://backingaus.innovation.gov.au/2004/research/qual_pub_res.htm
- Australian National University (ANU) 2003, 'Research and research training management report', *Research Office Research Data*. Retrieved 14 February 2005, from <http://www.anu.edu.au/ro/data/index.php>
- Australian Research Council (ARC). 2003, 'Implementation plan for national research priorities', *ARC Publications*. Retrieved 17 March 2005, from http://www.arc.gov.au/publications/arc_publications.htm#priorities
- Australian Science and Technology Council (ATSEC) 1989, *Profile of Australian Science: A study of the current state and potential of basic scientific research*. Canberra: Australian Government Publishing Service.
- Australian Vice-Chancellors Committee (AVCC) 2003, 'Evaluation of the knowledge and innovation reforms: AVCC submission', *Submission to Reviews & Inquiries 2003* Retrieved 11 March 2005, from <http://www.avcc.edu.au/documents/publications/stats/Research.pdf>
- Australian Vice-Chancellors Committee (AVCC) 2005, 'Higher education research data collection time-series data 1992-2003', *Research*. Retrieved 14 February 2005, from <http://www.avcc.edu.au/content.asp?page=/publications/stats/research.htm>
- AVCC/ACDP Working Party 1988, *Performance Indicators*. Canberra: Australian Vice-Chancellors Committee.
- Ball, R., & Wilkinson, R. 1994, 'The use and abuse of performance indicators in UK higher education', *Higher Education*, 27, 417-427.
- Bird, S. M., Cox, D., Farewell, V. T., Goldstein, H., Holt, T., & Smith, P. C. 2005, 'Performance indicators: Good, bad, and ugly', *Journal of the Royal Statistical Society Series A*, 168(1), 1-27.
- Bond University 2003, *Evaluation of the knowledge and innovation reforms reviewing Australia's block research funding scheme – executive summary*. Retrieved 10 February 2005, from http://www.dest.gov.au/highered/ki_reforms/submissions/pdf/k20.pdf

- Bordons, M., Fernandez, M. T., & Gomez, I. 2002, 'Advantages and limitations in the use of impact factor measures for the assessment of research performance', *Scientometrics*, 53(2), 195-206.
- Bourke, P., & Martin, B. 1992, 'Research: Gauging the width', *Australian Campus Review Weekly*, August 6-12, 11-12.
- Brennan, J. 1990, Quality assessment in the public sector in Great Britain. In L. C. J. Goedegebuure, P. A. M. Maassen & D. F. Westerheijen (Eds.), *Peer Review and Performance Indicators* (pp. 103-112). Utrecht: Lemma.
- Carotenuto, G., Lapegna, M., Zollo, G., Donato, A. D., & Nicolais, L. 2001, 'Evaluating research performance: The strategy of the University of Naples Federico II (Italy)', *Higher Education Policy*, 14, 75-90.
- Cave, M., Hanney, S., & Kogan, M. 1991, *The Use of Performance Indicators in Higher Education* (2 ed.). London: Jessica Kingsley.
- Cave, M., Hanney, S., Kogan, M., & Trevett, G. 1988, *The Use of Performance Indicators in Higher Education: A critical analysis of developing practice*. London: Jessica Kingsley Publishers.
- Cave, M., Kogan, M., & Hanney, S. 1990, The scope and effects of performance measurement in British higher education. In F. J. R. C. Dochy, M. S. R. Segers & W. H. F. W. Wynand (Eds.), *Management Information and Performance Indicators in Higher Education* (pp. 47-58). Maastricht: Van Gorcum.
- Charles, D., & Benneworth, P. 2002, 'Evaluating the regional contribution of an HEI: A benchmarking approach', *Publications 2002*. Retrieved 11 March 2004, from http://www.hefce.ac.uk/Pubs/hefce/2002/02_23.htm
- Chubin, D. E., & Hackett, E. J. 1990, *Peerless Science: Peer Review and U.S. Science Policy*. New York: SUNY Press.
- Cohen, D. 2004, 'NZ unis block rankings – research on the agenda', *Australian*, 24 March, p. 35.
- Council for the Humanities Arts and Social Sciences (CHASS). 2005, 'Measuring the quality and impact of research', *CHASS News and Media*. Retrieved 9 February 2005, from <http://www.chass.org.au/news/items/120105.html>
- Cuenin, S. 1987, 'The use of performance indications: An international survey', *International Journal of Institutional Management in Higher Education*, 11(2), 117-139.
- Department of Education Science and Training (DEST) 2000, 'Characteristics and performance indicators of Australian higher education institutions - 2000', *Higher Education*. Retrieved 4 March 2005, from <http://www.dest.gov.au/highered/statpubs.htm>
- Department of Education Science and Training (DEST) 2004a, 'Institutional Grants Scheme, 2004 guidelines'. Retrieved 10 February 2005 from <http://www.dest.gov.au/highered/research/igs.htm>
- Department of Education Science and Training (DEST) 2004b, 'Research Training Scheme, 2004 Guidelines'. Retrieved 10 February 2005, from <http://www.dest.gov.au/highered/research/rts.htm>
- Department of Education Science and Training (DEST) 2004c, 'Research Infrastructure Block Grants Scheme, Guidelines for 2004'. Retrieved 09 May 2005 from http://www.dest.gov.au/sectors/research_sector/programmes_funding/general_funding/research_infrastructure/research_infrastructure_block_grants_scheme.htm

- Department of Education Science and Training (DEST) 2005a, 'Research quality framework: Assessing the quality and impact of research in Australia - issues paper', *A Research Quality Framework: Issues Paper* Retrieved 11 April 2005, from http://www.dest.gov.au/resqual/issues_paper.htm
- Department of Education Science and Training (DEST) 2005b, 'Details of the government's response to the three reviews of the research system', *Higher Education* Retrieved 10 February 2005, from <http://www.dest.gov.au/highered/research/index.htm#how>
- Dochy, F., & Segers, M. (1990). *Selecting indicators on the basis of essential criteria and appropriate assessment methods for a quality assurance system*. Paper presented at the Paper prepared for the CHEMPs Conference 'Quality Assessment in Higher Education', March 16th, Utrecht, The Netherlands.
- Dochy, F. J. R. C., Segers, M. S. R., & Wijnen, H. F. W. 1990a, Preliminaries to the implementation of a quality assurance system based on management information and performance indicators: Results of a validity study. In F. J. R. C. Dochy, M. S. R. Segers & H. F. W. Wijnen (Eds.), *Management and Performance Indicators in Higher Education* (pp. 69-94). Maastricht: Van Gorcum, pp. 69-94.
- Dochy, F. J. R. C., Segers, M. S. R., & Wijnen, H. F. W. 1990b, Selecting performance indicators. In L. C. J. Goedegebuure, P. A. M. Maassen & D. F. Westerheijen (Eds.), *Peer Review and Performance Indicators* (pp. 135-154). Utrecht: Lemma.
- Doyle, K. 1995, 'Integrating performance measurement and resource planning in the university planning process', *Journal of Institutional Research in Australasia*, 4(1), 1-6.
- Elton, L. 1988, 'Appraisal and accountability in higher education', *Higher Education Quarterly*, 42(3), 208-229.
- Findlay, P. 1990, Developments in the performance indicator debate in the united kingdom. In L. C. J. Goedegebuure, P. A. M. Maassen & D. F. Westerheijen (Eds.), *Peer Review and Performance Indicators* (pp. 125-134). Utrecht: Lemma.
- Garlick, S., & Pryor, G. 2004, *Benchmarking the University: Learning about improvement*. Canberra: Department of Education Science and Training.
- Gilbert, R. 2005, 'The art of rating research', *Australian*, 26 January p. 26.
- Goedegebuure, L. C. J., Maassen, P. A. M., & Westerheijen, D. F. 1990, Developments in the performance indicator debate in the united kingdom. In L. C. J. Goedegebuure, P. A. M. Maassen & D. F. Westerheijen (Eds.), *Peer Review and Performance Indicators* (pp. 15-35). Utrecht: Lemma.
- Grigg, L., & Sheenhan, P. 1989, *Evaluating Research: The Role of Performance Indicators*. St Lucia: University of Queensland.
- Harris, G. T. 1993, *Research performance in Australian university economics departments*. Unpublished PhD Thesis, University of New England, Armidale.
- Hattie, J. 1990, 'Performance indicators in education', *Australian Journal of Education*, 34(3), 249-276.
- Illing, D. 2005, 'Scientist queries funds criteria' *Australian*, 14 January, p. 31.
- Irvine, J., & Martin, B. R. 1985, 'Basic research in the east and west: A comparison of the scientific performance of high-energy physics accelerators', *Social Studies in Science*, 15, 293-341.
- Jackson, N., & Lund, H. (Eds.) (2000). *Benchmarking for Higher Education*. Buckingham: Society for Research into Higher Education & Open University Press.
- Johnes, G. 1988, 'Research performance indications in the university sector', *Social Studies in Science*, 15(54-71).

- Johnes, J., & Taylor, J. 1990, *Performance Indicators in Higher Education*. The Society for Research into Higher Education: Open University Press.
- Johnson, S. B., & Osborne, R. D. 1997, 'Citation analysis and northern island: A quality measure?' *Higher Education Policy*, 10(3/4), 297-313.
- Joint Working Party on Higher Education Indicators 1994, 'Higher education performance indicators', *Diversity In Australian Higher Education Institutions, 1994* Retrieved 4 March 2005, from <http://www.admin.uwa.edu.au/stats/DEST/divers.htm>
- Jongbloed, B., & Westerheijen, D. F. 1994, Performance indicators and quality assessment in European higher education. In V. Borden & T. Banta (Eds.), *Using Performance Indicators to Guide Strategic Decision Making* (pp. 37-50). San Francisco: Jossey-Bass.
- Kells, H. R. 1990, Quality assessment in European higher education. In L. C. J. Goedegebuure, P. A. M. Maassen & D. F. Westerheijen (Eds.), *Peer Review and Performance Indicators* (pp. 211-217). Utrecht: Lemma.
- Kells, H. R. 1992, 'An analysis of the nature and recent development of performance indicators in higher education', *Higher Education Management*, 4(2), 131-138.
- Kells, H. R. 1993, *The Development of Performance Indicators for Higher Education*. Paris: OECD.
- Linke, R. D. 1991, *Performance Indicators in Higher Education*. Canberra: Australian Government Publishing Service.
- Linke, R. D. 1992, 'Some principles for application of performance indicators in higher education', *Higher Education Management*, 4(2), 194-203.
- Lucier, P. 1992, 'Performance indicators in higher education: Lowering the tension debate', *Higher Education Management*, 4(2), 204-214.
- Manaaki Whenua - Landcare Research 2004, 'Annual report 2004: For all dimensions of our performance', *Publications*. Retrieved 12 March 2005, from <http://www.landcareresearch.co.nz/publications/index.asp>
- Marginson, S. 1993, *Education and Public Policy in Australia*. Melbourne: Cambridge University Press.
- Martin, A. 2003, 'Universities quality agency: 2002 institutional audit reports, analysis and comment, ' *AUQA Occasional Publication*. Retrieved 9 March 2005, from <http://www.auqa.edu.au/qualityenhancement/occasionalpublications/index.shtml>
- McElwee, G. 1992, 'How useful are performance indicators in the polytechnic sector?' *Educational Management and Administration*, 20(3), 189-192.
- McKinnon, K. R., Walker, S. H., & Davis, D. 2000, *Benchmarking: A manual for Australian Universities*. Canberra: DETYA.
- McNay, I. 1997, 'The impact of the 1992 research assessment exercise in English universities', *Higher Education Review*, 29(2), 34-63.
- McNay, I. 1998, 'The Research Assessment Exercise (RAE) and after: 'you never know how it will all turn out'', *Perspectives: Policy and Practice in Higher Education*, 2(1), 19-22.
- Meek, V. L. 1995, 'Regulatory frameworks, market competition and the governance and management of higher education', *Australian Universities' Review*, 2-9.
- Naeve, G. 1988, 'On the cultivation of quality, efficiency and enterprise: An overview of recent trends in higher education in western Europe, 1986-1988', *European Journal of Education*, 23, 7-23.
- Politt, C. 1987, 'The politics of performance assessment: Lessons for higher education?' *Studies in Higher Education*, 12, 87-98.

- Politt, C. 1990, 'Measuring university performance: Never mind the quality, never mind the width', *Higher Education Quarterly*, 44, 60-81.
- Ramsden, P. 1999, 'Predicting institutional research performance from published indicators: A test of a classification of Australian university types', *Higher Education*, 37, 341-358.
- Reeve, R. C., & Hutchinson, P. J. 1988, 'The contribution of non-U.S. Institutions to academic accounting journals', *Abacus*, 24(1), 90-94.
- Research Assessment Exercise. 2001, 'A guide to the 2001 Research Assessment Exercise', *RAE 2001 - Publications* Retrieved 4 March 2005, from <http://www.hero.ac.uk/rae/Pubs/>
- Ronayne, J. 1984, *Science in Government*. Melbourne: Edward Arnold.
- Rudd, E. 1988, 'The evaluation of the quality of research', *Studies in Higher Education*, 13(1), 45-57.
- Sizer, J. 1988, 'The politics of performance assessment: Lessons for higher education? A comment', *Studies in Higher Education*, 13(1), 101-103.
- Sizer, J. 1990a, Funding councils and performance indicators in quality assessment in the united kingdom. In L. C. J. Goedegebuure, P. A. M. Maassen & D. F. Westerheijen (Eds.), *Peer Review and Performance Indicators* (pp. 155-182). Utrecht: Lemma.
- Sizer, J. 1990b, Performance indicators and the management of universities in the UK. In F. J. R. C. Dochy, M. S. R. Segers & H. F. W. Wijnen (Eds.), *Management Information and Performance Indicators in Higher Education* (pp. 5-46). Maastricht: Van Gorcum.
- Sizer, J. 1992, 'Performance indicators in government - higher institutions relationships: Lessons for government', *Higher Education Management*, 156-163.
- Sizer, J., Spee, A., & Bormans, R. 1992, 'The role of performance indicators in higher education', *Higher Education*, 24, 133-155.
- Slaughter, S., & Rhodes, G. 2003, Contested intellectual property: The role of the institution in united states higher education. In A. Amaral, V. L. Meek & I. M. Larsen (Eds.), *The Higher Education Managerial Revolution?* (pp. 203-228). Dordrecht: Kluwer Academic Publishers.
- Sornette, D., & Sajdenweber, D. 1999, 'Economic returns of research: The Pareto law and its implications', *The European Physical Journal B*, 8(2), 653-334.
- Stolte-Heiskanen, V. 1992, 'Research performance evaluation in the higher education sector', *Higher Education Management*, 4(2), 179-193.
- Tognolini, J., Adams, K., & Hattie, J. 1994, 'A methodology to choose performance indicators of research attainment in universities', *Australasian Journal of Education*, 38(2), 105-117.
- Tomlin, R. 1988, 'Research league tables: Is there a better way?' *Higher Education Quarterly*, 52(2), 204-220.
- Williams, R., & Van Dyke, N. 2004, 'The international standing of Australian universities'. Retrieved 11 April 2005, from http://www.melbourneinstitute.com/publications/reports/aus_uni/
- Wood, F. Q., & Meek, V. L. 1993, *Research Grant Management and Funding*. Canberra: ANUTECH.
- Yorke, M. 1995, 'Taking the odds-on chance: Using performance indicators in managing for the improvement of quality in higher education', *Tertiary Education and Management*, 1(1), 49-57.

APPENDIX 1: The McKinnon Benchmarks

SOURCE: McKinnon, Walker, and Davis (2000:vi-vii)

1. Governance, planning and management

- Benchmark: 1.1 Governance and leadership
- Benchmark: 1.2 University-wide planning
- Benchmark: 1.3 Strategic change initiatives
- Benchmark: 1.4 Equity planning
- Benchmark: 1.5 Clearly defined lines of responsibility & decision-making
- Benchmark: 1.6 Core business systems
- Benchmark: 1.7 Risk management
- Benchmark: 1.8 Teaching and research expenditure ratio
- Benchmark: 1.9 Corporate information systems
- Benchmark: 1.10 Organisational climate

2. External impact

- Benchmark: 2.1 Reputation
- Benchmark: 2.2 Competitiveness
- Benchmark: 2.3 Academic staff qualifications
- Benchmark: 2.4 Strategic community service
- Benchmark: 2.5 Exemplary community practices

3. Finance and physical infrastructure

- Benchmark: 3.1 Operating result
- Benchmark: 3.2 Diversity of revenue
- Benchmark: 3.3 Liquidity
- Benchmark: 3.4 External debt
- Benchmark: 3.5 Quick ratio
- Benchmark: 3.6 Academic salaries expenditure trends
- Benchmark: 3.7 Commercialisation: Net return on equity
- Benchmark: 3.8 Strategic asset management
- Benchmark: 3.9 Recurrent maintenance funding
- Benchmark: 3.10 Facilities maintenance backlog
- Benchmark: 3.11 Space management
- Benchmark: 3.12 Central teaching space usage and effectiveness
- Benchmark: 3.13 Large equipment utilisation
- Benchmark: 3.14 IT & T infrastructure

4. Learning and teaching

- Benchmark: 4.1 Learning and teaching plan
- Benchmark: 4.2 Course establishment processes
- Benchmark: 4.3 Scholarly teaching
- Benchmark: 4.4 Teaching environment
- Benchmark: 4.5 Effective academic review processes
- Benchmark: 4.6 Fitness of courses
- Benchmark: 4.7 Student progress ratio
- Benchmark: 4.8 First to second year retention trends
- Benchmark: 4.9 Equity quantitative success
- Benchmark: 4.10 Student satisfaction
- Benchmark: 4.11 Employability of Australian graduates

5. Student support

- Benchmark: 5.1 Student administrative services
- Benchmark: 5.2 Student services
- Benchmark: 5.3 Effectiveness of services

6. Research

Benchmark: 6.1 Research & research training planning

Benchmark: 6.2 Proportion of academic staff holding NCG OPS, or industry research grants

Benchmark: 6.3 Proportion of academic staff with direct involvement

Benchmark: 6.4 Research students' experience

Benchmark: 6.5 Research higher degree completion rates and times

Benchmark: 6.6 Research income trends

Benchmark: 6.7 Research higher degree completions per FTE academic staff

Benchmark: 6.8 Weighted research publications per FTE academic staff

Benchmark: 6.9 Impact of research

7. Library and information services

Benchmark: 7.1 Effectiveness of information planning processes

Benchmark: 7.2 Contributions to teaching and learning

Benchmark: 7.3 Provision of support for research

Benchmark: 7.4 Effectiveness of collaborative alliances

8. Internationalisation

Benchmark: 8.1 Internationalisation strategy

Benchmark: 8.2 Culture of internationalisation

Benchmark: 8.3. Balanced onshore international student programme

Benchmark: 8.4 Financing of the international student programme

Benchmark: 8.5 Students' exposure to international experience

Benchmark: 8.6 Management of offshore delivery

Benchmark: 8.7 Overseas links and activity

9. Staff

Benchmark: 9. 1 Strategic human resource planning

Benchmark: 9.2 Management of workforce

Benchmark: 9.3 Workforce diversity

Benchmark: 9.4 Career development/staff effectiveness

APPENDIX 2: Detail of Research Assessment examples

A2.1 Higher Education Research Data Collection - Australia

The indices they use in their collection are (AVCC 2005):

- University Research Income
 - Commonwealth Competitive Grants
 - Non-Commonwealth Competitive Grants
 - Total National Competitive Grants (Category 1)
 - Local Government
 - State Government
 - Other Commonwealth Government
 - Total Other Public Sector Funding (Category 2)
 - Australian Contracts
 - Australian Grants
 - Donations, Bequests and Foundations
 - International Funding
 - Syndicated Research & Development
 - Total Industry & Other Funding (Category 3)
 - Total Government Grants (Categories 1 & 2)
 - Category 4 - Cooperative Research Centre Funding
 - C'wealth Grants to CRCs
 - Non-university participants
 - Third party contributions
 - Total CRC funding
 - Total Financial Data (Categories 1, 2 and 3)
 - Total Financial Data (Categories 1, 2, 3, and 4)
- University Research Income: Financial Data from CRCs by Institution
 - Financial Data from CRCs
- University Research Publications
 - Books
 - Book Chapter
 - Journal Articles
 - Ref Work
 - Review Article
 - Conf Pub'n
 - Contract Report
 - Patent
- Research Training: Masters by Research and PhD completions
 - Masters by research
 - PhD by research

A2.2 Australian National University Research Quality Review 2004

The ANU prepared a portfolio of evidence as a Capabilities and Performance Statement for consideration by the Review Committee which included:

- *Comparative performance data for ANU against all other Australian universities;*
- *Evaluations by some 285 external peers of the quality of ANU research outputs by discipline and by academic unit;*
- *Contextual statements prepared by Deans and Directors for each University Centre, Faculty and Research School;*
- *Bibliometric analysis of the relative impact of ANU research by discipline;*
- *Surveys of domestic and international alumni of the ANU;*
- *A report of focus group discussions held with students and conducted by an independent survey organisation;*
- *A report of focus group discussions held with external stakeholders and conducted by an independent survey organisation;*
- *Results of graduate destination and satisfaction surveys over several years;*
- *Findings of internal surveys of students perceptions of ANU teaching over several years; and*
- *The 2004 report on progress in responding to the findings of the 2001 Review of ANU administration.*

For the evaluation of the quality of ANU's research outputs, all research active ANU staff were asked to identify their five best pieces of research, scholarship or creative work for the period 1995-2004. External peer assessors were then asked to rate the outputs in tranches of "international top 25 universities", top 50, top 100, and provide narrative judgements. International peer ratings were also sought to indicate the proportion of work judged to be in the top 5% and top 25% internationally.

Assessors were also provided with the Context Statement for the organisational unit which outlined the research interests and orientations of the area, along with comments on the research performance of the area and factors affecting performance. The Context Statements also contained a self-assessment rating on a five-point scale.

Source: ANU (2004 in DEST 2005a:54-55)

Issues relating to measuring excellence in research are also being taken up in a joint Australian National University (ANU) and DEST ARC Linkage Project which is exploring the use of quantitative indicators of research performance. The primary focus of the project is on the range of bibliometric indicators that can be used for assessing the written output of research. This work is expected to be completed by June 2005.

In addition, the Council for the Humanities, Arts and Social Science (CHASS) is undertaking a project to develop a set of possible measures suitable for assessing the

quality and impact of publicly funded Humanities, Arts and Social Sciences (HASS) research. This work is also expected to be completed by June 2005.

(DEST 2005a:55)

ANU Indices

As set out in ANU to 2005 Preparing Ourselves, the ANU is also implementing a range of measures of its performance under its objectives:

- Number of weighted publications per FTE academic staff (averaged over 3 calendar years). Based on DEST publication categories, ISI analysis, and appropriate output/impact measures for other disciplines;*
- Dollars per category of research grant. Other public sector research income, and industry and other research income per FTE. CRC and other Centre research income per FTE. Success rates per category of research grant. Ratio of total grant funding to FTE academic staff. Research income as proportion of total expenditure. Performance relative to other peer universities;*
- Trend data showing satisfaction levels for research degree students based on questionnaires;*
- Quality of supervision as revealed by PREQ;*
- PhD, Masters and Honours completions within time and by field of study;*
- HDR load per FTE academic staff;*
- Honours and distinctions accorded to staff;*
- Impact measures: citations, patents, licences. Major recognition through international/national invited lectures;*
- National and international linkages, service to government and professional societies;*
- Clearly communicated ANU research, research training, IP and technology transfer objectives, plans, and performance indicators reviewed annually by user groups and the Research Committee;*
- Annual growth of at least 10% in external resources including HDR student grants and consultancies; and*
- Increase in successful grant applications and bids for CRC's Centre of Excellence, major infrastructure, national and international initiatives, supporting bid managers where appropriate to enhance competitiveness.*

Quality assurance mechanisms for these measures are being developed as part of the ANU's Research and Research Training Management Plan.

(ANU 2003:3-4)

APPENDIX 3: The Charles and Benneworth Benchmarks

SOURCE: Charles and Benneworth (2002:12)

1 Enhancing regional framework conditions

Benchmark 1.1 Engagement in regional infrastructure planning and assessment

Benchmark 1.2 Investment in a high quality campus

Benchmark 1.3 Using HEI demand as lever to upgrade infrastructure

2 Human capital development processes

Benchmark 2.1 Access for students from disadvantaged groups

Benchmark 2.2 Retention of graduates in the region

Benchmark 2.3 Involvement in regional skills strategies

Benchmark 2.4 Responsiveness to regional labour market demands

Benchmark 2.5 Involvement of employers in developing the curriculum

Benchmark 2.6 Course provision for employers and employees

3 Business development processes

Benchmark 3.1 Strategic plan for business support

Benchmark 3.2 Creation of spin-off firms

Benchmark 3.3 Engagement in inward investment

Benchmark 3.4 Promoting graduate entrepreneurship

Benchmark 3.5 Graduate start-ups arising from HEI programmes

Benchmark 3.6 Availability of entrepreneurship modules

Benchmark 3.7 Student placements with local employers

Benchmark 3.8 Incentives for staff to engage with business

4 Interactive learning and social capital development processes

Benchmark 4.1 Involvement in regional governance

Benchmark 4.2 Contribution to regional economic analysis

Benchmark 4.3 Analysis of regional futures

Benchmark 4.4 Staff exchanges

5 Redistributive processes

Benchmark 5.1 Contributing to healthy cities and health promotion

Benchmark 5.2 Support for community-based regeneration

Benchmark 5.3 Student community action

6 Cultural development

Benchmark 6.1 Cultural strategy

Benchmark 6.2 Provision of cultural facilities

Benchmark 6.3 Impact on local tourism

Benchmark 6.4 Levels of participation by the community

Benchmark 6.5 Fostering regional cultural identities

Benchmark 6.6 HEI spin-offs to the cultural sector

7 Promoting sustainability

Benchmark 7.1 Sustainability at the heart of HEI governance

Benchmark 7.2 Promoting sustainability through the curriculum

Benchmark 7.3 Performance against environmental management systems