

## Little fish in a big pond: Towards research performance metrics for smaller institutions

SHARON KITT<sup>1</sup> AND JONATHAN WEARNE<sup>2</sup>

<sup>1</sup>*Robert HT Smith Research Fellow, University of Ballarat*

<sup>2</sup>*Director, Collaborative Research Network, University of Ballarat*

Submitted to the *Journal of Institutional Research* January 20 2013, accepted for publication

April 17 2013.

---

---

### Abstract

Each year, institutions eagerly await reports from *Shanghai Jiao Tong University*, *Times Higher*, *QS*, and other organisations that create and publish international rankings of university performance. The metrics included in league tables and rankings—research income, research staff, number of doctoral candidates, numbers of publications—are common to other measures of research performance. Invariably, these ‘four pillars’ of research performance measurement are used as proxy measures of quality, but they are in fact quantity measures, reflecting that size does matter. For smaller and regional institutions that are not listed in the Top 100, or not even players in the Top 500, it is difficult to demonstrate and measure quality when quantity is such a factor. This article examines the history of the research performance measurement within Australian higher education, and questions the current validity and focus of these metrics. It further explores the context of these metrics, and considers the requirements for ‘little fish’ in the higher education ‘pond’ to demonstrate excellence in research.

**Keywords:** research performance, new public management, indicators, regional Australia

---

The last decade has seen a proliferation of the measurement of research productivity within the higher education sector, not just within Australia but worldwide. While teaching and learning may be the public interface of the university, servicing hundreds of thousands of Australian and international students, it is performance in research that drives much of the funding. Research performance measures are used as a proxy for the reputation and performance of the institution within local and international contexts. Indeed, the second half of each calendar year is now dubbed ‘rankings season’, as it sees the release of a range of international higher education performance assessments in the form of league tables or rankings. For the smaller higher education institutions within Australia, typically located in the regional areas, the rankings season rarely features the work of their institutions. Smaller

---

This article was first presented at the Annual Conference of the Australasian Association for Institutional Research, *Evolution of IR practice and use of IR*, Terrigal, New South Wales, November 12–14, 2013.

*Correspondence to:* Sharon Kitt, Robert HT Smith Research Fellow (Rural and Regional Research) at University of Ballarat. E-mail: s.kitt@ballarat.edu.au

and/or regional institutions find it difficult to compete with the research size and capacity of the ranked universities that so often influence position on these rankings.

Considering the current invisibility of smaller Australian institutions within the world rankings, this subset of the higher education sector must be given consideration regarding demonstration of value. Rather than opting out or disparaging the rankings, are there ways in which smaller institutions can demonstrate their worth that do not rely merely on the size of the institution? If so, what are the theoretical constructs that underlie the development of such metrics, and how can the little fish in the big higher education pond capture this in new or revitalised indicators?

### **Measurement and the Higher Education Sector**

The drivers of performance measurement derive from the management and administrative science theories popularised in the 1980s that continue to have influence over measurement policies and strategies today. Perhaps the most influential of these in the higher education sector is New Public Management (NPM). This management theory has as one of its central tenets the requirement for demonstrable accountability for performance. Thus, NPM espoused a clear and dominant focus on results (O’Flynn, 2007) and, as such, organisational performance became not just a conceptual ideal for the private sector, but public sector organisations were also expected to be productive and accountable businesses.

The concepts of NPM were catapulted into the hearts and minds of Australian university management in 1987 through the Dawkins reforms and its ‘emphasis on efficiency and quality’ (Dollery, Murray, & Crase, 2006, p. 91), reflected in an increased emphasis on benchmarks, performance indicators and monitoring in all aspects of research, teaching and learning (Bleiklie, Enders, Lepori, & Musselin, 2011). From the Dawkins reforms sprang the major research performance indicators still used today that underpin much of current research reporting—the ‘four pillars’ of research student load, research student completions, research income, research publications.

Simultaneously with the Dawkins reforms there arose an emphasis on internal performance monitoring by senior executives of organisations (public or private), again through the influence of NPM. There was much enthusiasm about the potential contribution the balanced scorecard approach could make to understanding and intervening in the research health of universities. The four pillars of research performance were viewed as the ideal metrics for these scorecards: they consisted of numeric and relatively accessible data; and they were brought together for university executives in a ‘single management report, many of the seemingly disparate elements of a . . . competitive [research] agenda’ (Kaplan & Norton, 1992, p. 73). This enthusiasm regarding balanced scorecards spilled over into the academic domain (Neely, Gregory, & Platts, 2005), with the report by Kaplan and Norton (1992) the most cited for eight out of the 10 years prior to 2005 in the field of performance measurement—but it appears to have made no impact on the field of *research* performance measurement. To this day, the four pillars of data continue to be captured and reported on an annual basis, and they are still presented as a collection of ‘research statistics’ to represent research activity, even though the actual research work leading to the publication of a paper or the receipt of funding may have occurred at a significantly earlier point in time.

### The Four Pillars in Context

In terms of Commonwealth funding based on research performance, a block grant system uses the four pillars of research performance in various formulae to annually allocate \$1.63 billion (Department of Industry, Innovation, Science, Research and Tertiary Education [DIISRTE], 2012) to institutions to use at their discretion for research and research training. To recognise the differing costs incurred with research production, a weighting is applied to higher degree completions, alleviating the expenses incurred by high-cost research programs (medicine, engineering). There are no other modifiers to the data used, with the larger institutions who report the largest quantitative value of indicator inputs consequently having returned to them a larger portion of the block grant ‘pie’. In summary, quantity and size are rewarded proportionally by the mechanisms of the funding formulae—inputs = outputs.

Australian universities have been required to submit figures summarising research performance since 1992 via the Higher Education Research Data Collection (HERDC). During this time, other than some minor refinements (adding and subtracting subcategories of data, minor definitional changes), there have been few changes to the scope of main indicators. While some elements of quality assessment are built into the types of data collected (peer-reviewed income categories and peer-reviewed publications), it is not an explicit aim of these data collection exercises to judge the overall quality of research inputs and outputs of an institution. Additionally, there are no moderators based on the size of the staffing cohort, institutional age or the relative institutional or discipline context within which the research was conducted. Without consideration or at the very least acknowledgement of the impact of these moderators, comparison between and across institutions may well be flawed.

### Individual Performance Versus Organisational Performance

The debate on development of performance measures often deals with the micro level of performance, the indicators themselves, and these are then applied to the individuals that create the works of research, the academic staff. Even where research performance models have been conceptualised, they start from the point of the individual (Bazeley, 2010), and are then summed up to the organisational level, rather than using the needs of the organisation as the starting point itself. This makes the discovery or creation of relevant metrics for smaller institutions difficult, as issues surrounding the validity of the metrics themselves within the context of the environment they are produced within are given scant attention within the literature. Neely’s 2005 *The Evolution of Performance Measurement Research* is an update on his 1994 literature review and an important way-point in the emerging field of academic inquiry on performance measurement. Neely found several distinct phases in the pre-2005 literature. Two of these provide an intriguing background to the performance measurement in academia. In the 1980s, the emphasis was on the problems of performance measurement, and their tendency to result in short-term and dysfunctional consequences. In the 1990s, when the Dawkins reforms and the quantitative pillars of research reporting were taking hold in Australian universities, the international literature had moved decisively towards finding potential solutions to the problems of measurement, and a search for frameworks that might provide useful ways of addressing well-documented issues. While the business and not-for-profit sectors made this move, a particular form of performance measurement—the citation counts and indices tallied into league tables described earlier in this article—appears to have taken deeper root in government funding formula and university business. The accountability campaign within the Australian higher education and public service appears to be the driver

for this particular form of performance accounting (Taylor, 2009). Rarely are reflections on the establishment and application of performance measure indicators for universities published; such assessment usually relies on the balanced scorecard approach or other tallying principles (Chen, Wang, & Yang, 2009; Philbin, 2011; Phusavat, Ketsarapong, Ranjan, & Lin, 2011), rather than any true assessment of the ontological justification of research performance measurement

One common analogue for organisational research performance is citation metrics, which actually measure the frequency an individual researcher or paper has been cited in the literature. Interestingly, when first introduced to the higher education sector, bibliometric (citation) data was also seen as a being highly useful at the policy and strategy level. One of the earliest descriptors of bibliometric data analysis published nearly thirty years ago envisaged bibliometrics as:

A “monitoring device” for research management and science policy. It enables research policy-makers to ask relevant questions in order to find an explanation of the bibliometric results in terms of policy relevant factors. This offers them the possibility of obtaining relevant information needed to make justified policy decisions. (Moed, Burger, Frankfort, & Van Raan, 1985, p. 147)

Nowadays, citation metrics have little association with institutions; however, over the last ten years their use to assess research has quadrupled.<sup>1</sup> This positivist focus on numbers as equating to performance and thus being interpreted as a corollary to institutional value has led to the “‘metricization’ of the academy” (Burrows, 2012, p. 356). Indeed, so prevalent is the focus on individual-level analytics that most researchers now quote their citation metrics—such as h-index value or equivalent (Froghi et al., 2012)—in funding and promotion applications.

Recent research performance exercises have begun to focus on quality metrics as well as the traditional and well-collected quantity metrics. The *Excellence in Research for Australia* (ERA) initiative ‘will evaluate the quality of the research undertaken in eligible higher education providers’ (Australian Research Council, 2011, p. 9). Some of the same data that are reported for the regular annual data collections are used for ERA (research income, publications), but with an emphasis on quality through additional peer review or citation analyses. Although the focus on quality has been welcomed by the higher education sector, the ERA evaluation and data are still hostage to the size factor: as part of its evaluation process, volume-activity analyses will be undertaken ‘on the basis of total research outputs, research income and other research items’ (Australian Research Council, 2011, p. 9). Additionally, the funding that stems from ERA performance—a measure based on quality—will also only be available to institutions that meet or exceed a particular research income threshold, a measure based purely on quantity (DIISRTE, 2012a, p. 3).

For the 2012 allocation of *Sustainable Research Excellence* (SRE) funding, data provided by DIISRTE demonstrates that twelve institutions were not recipients of SRE Threshold 2 performance-based funding as they did not meet the research income threshold (DIISRTE 2012a)—notably, these also happen to be the smallest or the most geographically isolated institutions in Australia.

---

<sup>1</sup> Scopus search conducted September 5 2012: citation metrics, citations, bibliometric, or citation analysis articles published between 2001 and 2011.

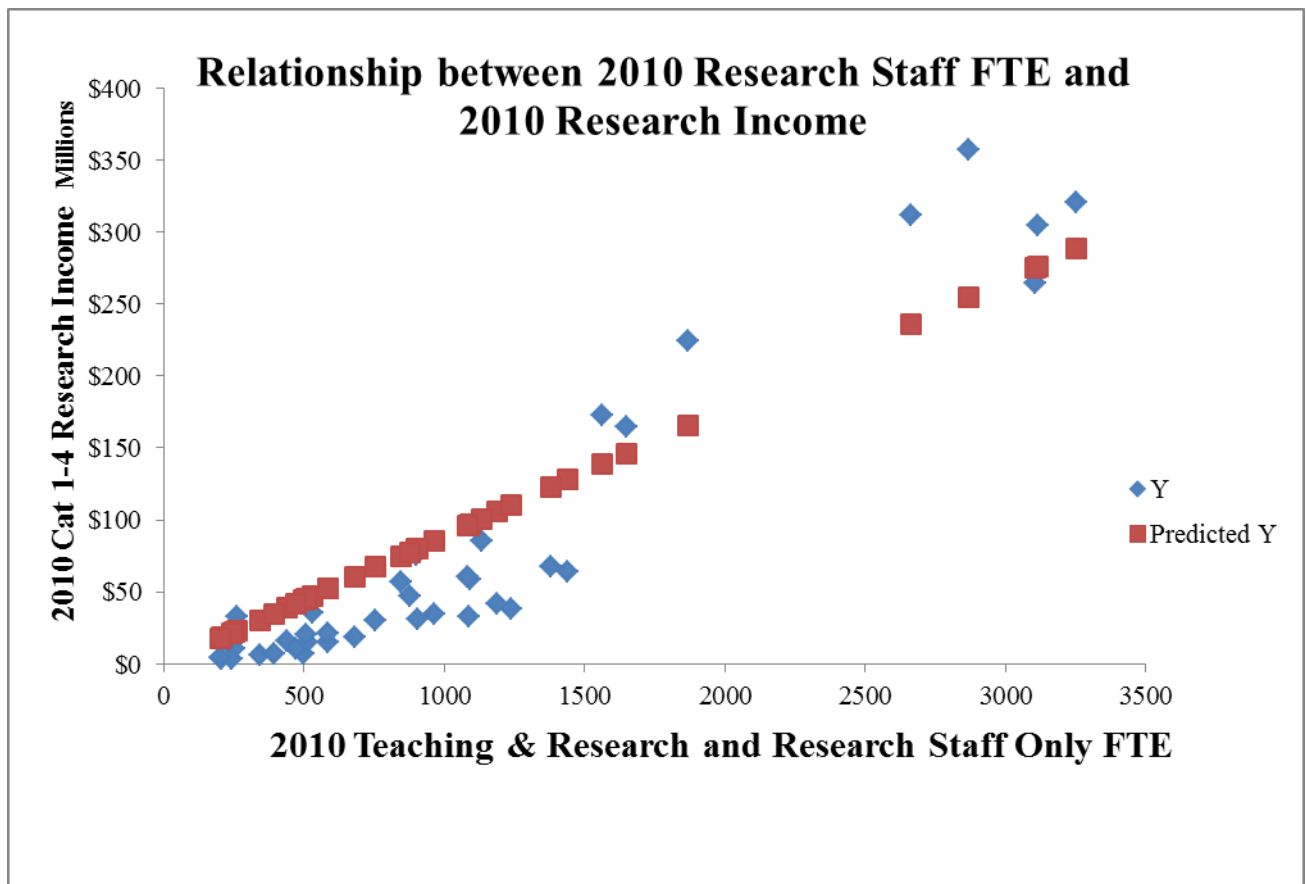


Figure 1. Correlation between research income and staff size.

A basic analysis shows that size of the institution and the size of its research income are very highly correlated (Figure 1) ( $r = .90$ ,  $p < .05$ ). Thus, smaller institutions with a smaller staffing cohort are likely to have a smaller research income base, and then smaller returns on their research investment, leading to smaller levels of reinvestment in research and research training, reflected in smaller research outputs, and so on.. That is, smaller institutions struggle to demonstrate the value of their research through purely quantitative measures and predictive modelling suggests they will always struggle due to size. Accordingly, through two of the main funding mechanisms for rewarding research performance in the Australian higher education sector (ERA and HERDC), the impact of size on performance rewards or recognition is of great concern. Smaller institutions must rise to the challenge to demonstrate their relative value to the sector within a contextual environment that sees organisational size as a critical component of reward.

In 2011 a group of rural and regional institutions formed the Regional Universities Network (RUN). With six foundation universities, RUN was established as a regional lobby group to governments, as well as a vehicle to promote internetwork collaboration. As can be seen from Figure 2, these universities have a small research output when compared to the rest of Australia, even when the universities' research performance results were combined (DIISRTE, 2012b).

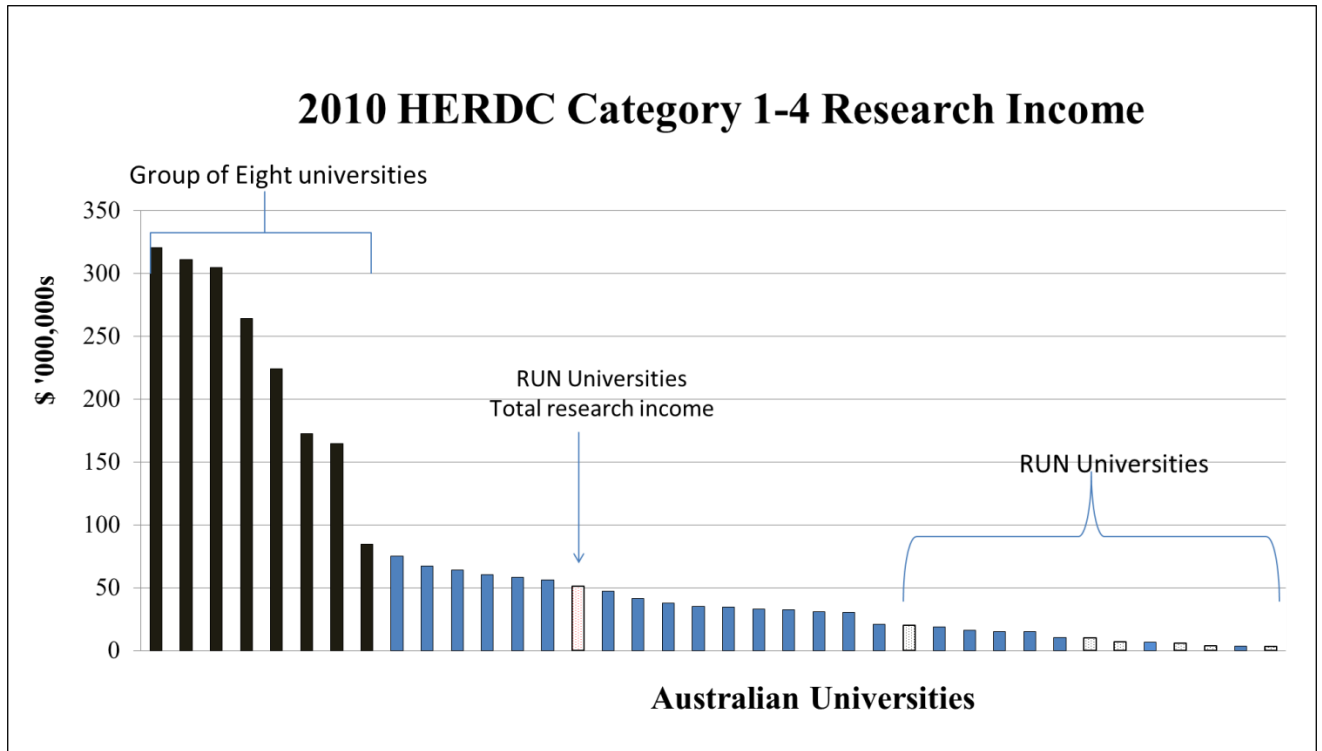


Figure 2 2010 HERDC research income for all Australian universities.

Regardless of the quantity of research achieved at these institutions, the above figure demonstrates little regarding the quality of research performance; rather it merely demonstrates that these institutions have a smaller research income in relation to their sector counterparts. Other smaller and regional institutions have opted out of the league table and rankings arms race all together. In 2012 the Vice-Chancellor of James Cook University decided not to participate in one of the major ranking exercises as

the strengths and virtues of more specialised, smaller, more locally oriented universities don't translate into a meaningful rank order position, not because they are not doing good and important work but simply because they are more specialised, smaller or locally oriented. (Harding, 2012)

Colyvas and Powell (2009) also suggest that it is doubtful that 'existing metrics focused on magnitude' are that useful, whereas the development of framework of 'more local, contextualized indicators can be harnessed' (Colyvas & Powell, 2009, p. 83). It is of no surprise then that the authors of this article have been tasked to examine ways in which research performance can be demonstrated in a manner that is not interdependent on the size of the institutions.

### Key Questions to Consider

Taking all of the above into consideration, one primary difficulty that arose was to separate the goal of the developing a relevant framework for smaller institutions without allowing currently available data to drive the design.. This has been reported within the literature as a common issue, where 'indicator developers will tend to concentrate (first) on developing indicators of those things that are easiest to measure, which may not be the variables most pertinent to STI [science and technology indicators] policy or management' (Freeman & Soete, 2009, p. 583). In recognising this, it became apparent there was a need to

move the development phase back into the theory and literature surrounding performance. It has been important to unearth or rediscover the cogent reasons behind indicators to ensure their actual fit for purpose within smaller institutions. In essence, the work has shifted from the manufacturing of a raft of new indicators separated from theory, towards a strong theoretical foundation upon which future development of specific research performance metrics could be based. In support of this are Kerssens-van Drongelen and Cooke (1997), who encouraged the early developers of metrics to ‘not unthinkingly copy the concepts proposed by others but design their own tailor-made system, suiting the purpose(s) of measurement and the peculiarities of their R&D setting’ (Kerssens-van Drongelen & Cooke, 1997, p. 356).

Regardless of the dearth of epistemological examination of the role and types of research performance measurement within universities, there still exists within the body of literature as a whole several interlocking questions that surround the production of performance metrics and performance frameworks. The literature asks that the following be considered:

- Does it review (Chen et al., 2009) or reward (Van Veen-Dirks, 2010) past performance?
- Does it drive or guide future performance (Franco-Santos, Lucianetti, & Bourne, 2012) through supporting strategic or organisational goal-setting (Micheli & Manzoni, 2010)?
- Does it build (positive) reputation (Boyd, Bergh, & Ketchen, 2009)?
- Does it work to influence policy (Nelson, 2012)?
- Does it serve an external accountability or compliance need (Taylor, 2009)?

This list above is by no means an exhaustive one. Furthermore, these points are not a checklist of attributes that research performance metrics or systems should exhibit; rather, they are but conversation starters regarding the contextualisation of a metric within its geographical and philosophical environment. For example, the role of metrics in being able to build (or break) reputation and to influence policy has been given scant attention in the research metrics space. The contextualisation of the metric within its environment must be considered as valuable as the numeric value of the metric itself. For smaller institutions, the regional isolation and community engagement themes that are often at the core of their vision or mission statements must be considered and must be truly integrated into any assessment of performance.

While universities have a tradition of peer review of scholarly work that rewards quality, the authors are left with the impression that research into research performance measurement in higher education is somewhat unhinged from approaches and the lessons in other settings (Agostino & Arnaboldi, 2012). It appears that academe was somehow left behind in the universities’ clamour to meet compliance reporting requirements. These requirements, a spill-over from the business sector into accountability initiatives in the public services, have been implemented within universities and exhibit a parlous intellectual state. The best the academe offers on research performance measurement seems to be a retrospective gaze at the ‘impacts’; usually once the paradigm is implemented to line-manage individual performance and generating perverse outcomes (see for example Bogt & Scapens, 2012).

A more comprehensive study is required to test the contention that universities, both big and small, were as systematically and intellectually unprepared for the risks and the opportunities brought by ‘new public management’ and organisational theory, as this preliminary survey of the literature suggests. An academe unprepared for undertheorised measurement tradition imported by a public service ushering in a new era of accountability is certainly an irony worth exploring.

### **Conclusion**

Instead of systematically exploring how to use performance measures to assist research institutes and university departments devise achievable goals, track progress, and align effort of researchers, it appears that universities have participated in a massive research performance data collection juggernaut that is scaled up in meta-metric analysis and benchmarking. Instead of research institutions deciding what is to be or what has been achieved, the noise of citations and other metrics that serve benchmarking, meta-analysis and university league tables cloud and crowd discussion in this space. The smaller institutions find it difficult to both raise and discuss these issues within the Australian university landscape, as their size and relative lobbying power is dwarfed against the sheer might of the large metropolitan universities and their representation via peak groups such as the Group of Eight.

With this in mind, further work must be done in the area of relevant performance measures with an eye to the future requirements of large and small institutions. The next pressing research question is how to enrich the scholarly analysis and relevance of measurement in university research management. A one-size-fits-all approach is patently disadvantageous to smaller institutions. A good beginning would be to fill the hole identified in this article and by others (Bogt & Scapens, 2012; Butler, 2010), and to tackle the apparent lack of research on the nature and consequences of performance measurement in universities.

Discussion is starting to surface within the sector surrounding impact measures within research performance and evaluation. This, in turn, is almost predicted by the emergence of the public value paradigm, where there is a shift from results to relationships (Taylor, 2009). Rather than focus on the measures associated with impact, further work needs to borrow from the numerous critiques of performance management/measurement in the public sector and business community, reflecting those against the impact of research produced among the unique characteristics of smaller and regional universities. From there it may be possible to design and implement research measurement approaches that support long-term institutional development, and provide a counter to the bibliometric mania that diminishes the contribution a small fish can make in a big pond.

### **Acknowledgments**

The authors acknowledge the support of the Collaborative Research Network for Self-Sustaining Research in the production of this article.

### **References**

Agostino, D., & Arnaboldi, M. (2012). Design issues in balanced scorecards: The “what” and “how” of control. *European Management Journal*, 30, 327–339.



- Australian Research Council. (2011). *ERA 2012 Submission Guidelines*. Canberra ACT: Commonwealth of Australia. Retrieved from [http://www.arc.gov.au/pdf/era12/ERA2012\\_SubmissionGuidelines.pdf](http://www.arc.gov.au/pdf/era12/ERA2012_SubmissionGuidelines.pdf)
- Bazeley, P. (2010). Conceptualising research performance. *Studies in Higher Education*, 35(8), 889–903. doi:10.1080/03075070903348404
- Bleiklie, I., Enders, J., Lepori, B., & Musselin, C. (2011). NPM, network governance, and the university as a changing professional organization. In T. Christensen & P. Laegreid (Eds.), *The Ashgate Research Companion to New Public Management* (pp. 161–176). Farnham, United Kingdom: Ashgate Publishing Limited.
- Boyd, B.K., Bergh, D.D., & Ketchen, D.J. (2009). Reconsidering the reputation—performance relationship: A resource-based view. *Journal of Management*, 36(3), 588–609. doi:10.1177/0149206308328507
- Burrows, R. (2012). Living with the h-index? Metric assemblages in the contemporary academy. *The Sociological Review*, 60(2), 355–372. doi:10.1111/j.1467-954X.2012.02077.x
- Butler, L. (2010). Impacts of performance-based research funding systems: A review of the concerns and the evidence. In OECD, *Performance-based Funding for Public Research in Tertiary Education Institutions* (pp. 127–165), Paris: OECD Publishing. Retrieved from <http://www.oecd.org/science/scienceandtechnologypolicy/performance-basedfundingforpublicresearchintertiaryeducationinstitutionsworkshopproceedings.htm>
- Chen, S.-H., Wang, H.-H., & Yang, K.-J. (2009). Establishment and application of performance measure indicators for universities. *The TQM Journal*, 21(3), 220–235. doi:10.1108/17542730910953004
- Colyvas, J.A., & Powell, W.W. (2009). Measures, metrics, and myopia: The challenges and ramifications of sustaining academic entrepreneurship. In G.D. Libecap (Ed.), *Measuring the social value of innovation: A link in the university technology transfer and entrepreneurship equation* (Vol. 19). Bingley: Emerald Group Publishing.
- Department of Industry, Innovation, Science, Research and Tertiary Education. (2012). *The process for determining sustainable research excellence (SRE) scheme grant amounts—2012 allocations*. Canberra, Australia. Retrieved from <http://www.innovation.gov.au/Research/ResearchBlockGrants/Documents/SREProcessCalculations2012.pdf>
- Dollery, B., Murray, D., & Crase, L. (2006). Knaves or knights, pawns or queens? An evaluation of Australian higher education reform policy. *Journal of Educational Administration*, 44(1), 86–97. doi:10.1108/09578230610642674
- Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2006). New Public Management is Dead: Long Live Digital Era Governance. *Journal of Public Administration Research and Theory*, 16, 467–494.
- Franco-Santos, M., Lucianetti, L., & Bourne, M. (2012). Contemporary performance measurement systems: A review of their consequences and a framework for research. *Management Accounting Research*, 23(2), 79–119. doi:10.1016/j.mar.2012.04.001

- Freeman, C., & Soete, L. (2009). Developing science, technology and innovation indicators: What we can learn from the past. *Research Policy*, 38(4), 583–589. doi:10.1016/j.respol.2009.01.018
- Froghi, S., Ahmed, K., Finch, A., Fitzpatrick, J.M., Khan, M.S., & Dasgupta, P. (2012). Indicators for research performance evaluation: An overview. *BJU International*, 109(3), 321–324. doi:10.1111/j.1464-410X.2011.10856
- Harding, S. (2012, May 23). Why James Cook chose to become a rank outsider. *The Australian*. Retrieved from <http://www.theaustralian.com.au/higher-education/opinion/why-james-cook-chose-to-become-a-rank-outsider/story-e6frgcko-1226363850473>
- Kaplan, R.S., & Norton, D.P. (1992). The balanced scorecard: Measures that drive performance. *Harvard Business Review*, 70(1), 71–9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10119714>
- Kerssens-van Drongelen, I.C., & Cooke, A. (1997). Design principles for the development of measurement systems for research and development processes. *R&D Management*, 27(4), 345–357. doi:10.1111/1467-9310.00070
- Micheli, P., & Manzoni, J.-F. (2010). Strategic performance measurement: Benefits, limitations and paradoxes. *Long Range Planning*, 43(4), 465–476. doi:10.1016/j.lrp.2009.12.004
- Moed, H.F., Burger, W.J.M., Frankfort, J.G., & Van Raan, A.F.J. (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy*, 14(3), 131–149. doi:10.1016/0048-7333(85)90012-5
- Neely, A., Gregory, M., & Platts, K. (2005). Performance measurement system design: A literature review and research agenda. *International Journal of Operations & Production Management*, 25(12), 1228–1263. doi:10.1108/01443570510633639
- Nelson, A.J. (2012). Putting university research in context: Assessing alternative measures of production and diffusion at Stanford. *Research Policy*, 41, 678–691. doi:10.1016/j.respol.2011.11.004
- O'Flynn, J. (2007). From new public management to public value: Paradigmatic change and managerial implications. *Australian Journal of Public Administration*, 66(3), 353–366. doi:10.1111/j.1467-8500.2007.00545.x
- Philbin, S.P. (2011). Design and implementation of the balanced scorecard at a university institute. *Measuring Business Excellence*, 15(3), 34–45. doi:10.1108/13683041111161148
- Phusavat, K., Ketsarapong, S., Ranjan, J., & Lin, B. (2011). Developing a university classification model from performance indicators. *Performance Measurement and Metrics*, 12(3), 183–213. doi:10.1108/14678041111196668
- Taylor, J. (2009). Strengthening the link between performance measurement and decision making. *Public Administration*, 87(4), 853–871. doi:10.1111/j.1467-9299.2009.01788.x
- ter Bogt, H.J., & Scapens, R. (2012). Performance management in universities: Effects of the transition to more quantitative measurement systems. *European Accounting Review*, 21(3), 451–497. doi:<http://dx.doi.org/10.1080/09638180.2012.668323>

Van Veen-Dirks, P. (2010). Different uses of performance measures: The evaluation versus reward of production managers. *Accounting, Organizations and Society*, 35(2), 141–164. doi:10.1016/j.aos.2009.02.002