

University of Warwick
Department of Economics

QR and Metrics

Prepared by Mark Harrison and Michael Waterson

This version 30 June 2006; first draft 26 June 2006.

Summary. This note presents strong arguments for the retention of an element of peer review of research output in any future model for the allocation of QR funding. Any metric based mainly on external research funding is flawed, and we consider the implications. Measures of influence such as citation rates and journal impact factors have some advantages, but an element of peer review remains essential.

Introduction

The government is looking for a new way of allocating QR funding among universities and is consulting on the replacement of a peer-review based metric, the RAE, by a metric based on external research funding.¹ There is evidence that the RAE's competitive focus on relative research quality has been generally productive for average quality in the UK across institutions over several cycles. But the criticisms of the RAE are significant. In particular, it is said, the RAE is backward-looking and infrequent and so cannot measure research quality in a responsive, forward-looking way; also that peer review is hard but not impossible to manipulate, especially by short-term hiring strategies focused on infrequent census dates.²

Coming from a leading social-science department, we are glad to seize the chance to move away from the more obvious distortions arising from the RAE framework. At the same time we wish to warn against the idea that a metrics-based approach should focus exclusively on the volume of competitive research funding.

The Purpose of Funding Allocation

We should first ask: what is the fundamental purpose of QR funding allocation? The DfES consultation document is surprisingly coy. It does refer, however, to a need to find "the best or the most efficient way to allocate resources while driving up research quality" (para. 2.9). More succinctly, it seems, *they wish to reward efficiency in an efficient way.*

Evidently, some institutions produce high quality research more efficiently than others. If this were not so, the allocation of QR funding would not matter because, in the absence of efficiency differences, a given funding amount would produce the same research quality wherever it was applied. When efficiency varies, however, shifting QR funding towards the more efficient institutions must have the *immediate* effect of improving research productivity on average. There should also be a *long-run* effect, since shifting QR funding

¹ DfES (13 June 2006), *Reform of Higher Education Research Assessment and Funding.*

² There is also a perception that the RAE is costly, but the evidence tends to show that its direct costs are smaller in proportion to the funding allocated than those of the Research Councils, for example.

towards the more efficient institutions rewards research efficiency and must sharpen the incentives for institutions to improve efficiency over time.

This suggests that the fundamental purpose of any metric used to allocate QR funding should be to identify *research efficiency* – the efficiency with which research of given quality is produced. This is true whatever dimensions of quality are chosen. It also should identify efficiency currently, rather than far in the past – an obvious defect of the RAE as it is now.

External Funding

The initial Treasury documentation showed that there is a high bivariate correlation between competitive funding-based metrics and RAE-based QR funding. This suggests that the Research Councils have indeed recently allocated competitive funding in some proportion to *past research quality*. (It is not clear whether they would do this so well in the absence of past RAE scores to guide them. But that is not the point.) The point is this: does the recent allocation of external funding measure *current research efficiency*?

We will make four points. Reducing them to their simplest: if external research funding alone were a good predictor of research success, by implication it would not matter to whom it was given.

(1) There is an obvious distinction between peer review of research *output* and a metric based on external funding. External research funding is an *input* to research; measuring efficiency, or output per unit input, by a measure of inputs must be wrong in principle. It is like measuring research performance by the size of a department's faculty. It is *possible* that a large department will do good research. But whether it will *actually* do so depends entirely on the quality of the people in it and the quality of their interaction. Knowing their number tells us nothing about this. MIT, for example, unquestionably produces research of world quality across many fields, for example, but has fewer than 1,000 faculty members across its roughly 35 divisions.

(2) Of course, it is possible that a large department became large by being good at research in the past. That must be the argument for using external research funding as the basis of a metric or an element in the metric. Implicitly, it must be thought that departments and universities have recently attracted funding by being good in the past. Note the implication, however: a funding metric will be *more* backward looking than a peer-review metric, since peer review in period t must be based on research output in period $t - 1$, while external funding in period t tends to be based on peer review in period $t - 1$, which in turn is based on research output in $t - 2$.

(3) Much of the DfES consultation is taken up with alternative models for calibrating research funding to different discipline groups: STEM, the social sciences, and arts, for example. The underlying reason is that research technologies differ across disciplines; some are intrinsically more input-intensive than others. Without moderation for this effect, a metric based on external funding would unduly motivate institutions to specialise in the funding-intensive physical sciences and medicine. Put differently, if £1M of external funding buys far less research *output* in astronomy than it does in mathematics, a bias towards astronomy cannot be right. All the consultation about alternative models is no more than an effort to correct for such biases that are intrinsic to the method.

(4) Finally, external funding is intrinsically more manipulable than research quality itself. It is a simple point. If the Research Councils do their job well, the gaining of external funding must be a function of two variables, past research quality and the effort put into bids for funds. Suppose that there is some degree of substitutability between these two things so that, conditional on higher past research quality, success in a given bid can be obtained with lower effort. Then, if QR funding is made to depend on external funding, it will benefit poorer quality institutions to put higher effort into getting external funding and thus obtain QR funding out of proportion to their quality.

Bibliometrics

Bibliometrics have the great advantage that they measure output, not inputs. They seek to capture the influence of a piece of work, which is an important dimension of research quality – in some ways, the only one that matters. But they suffer from well known difficulties. The influence of a piece of work is approximated by the citations that it accumulates, but the accumulation of citations takes time and judging true influence takes years. Citation rates fail the forward-looking test just as badly as external funding.

For current work, an alternative is the impact factor of the journal in which it is published. We see that a colleague has her first paper accepted in an international journal and we judge her research potential by this fact. In short, we transfer responsibility for the judgement to the editor and the peer review process, which confers the average reputation of the articles the journal has published in the past to current work. Thus, the true information we have gained is again backward looking. It is also noisy, since journals make mistakes, at least by comparison with subsequent citations; in economics, for example, the citation rates of the most influential papers in middle and low-ranking journals overlap substantially the citations of less influential papers in top journals (Oswald 2006).

Finally, just as disciplines differ in funding intensity, they differ also in size and scope, and this alone creates difficulties of comparing citations and impacts across disciplines equivalent to those arising from external funding measures. Additionally, citation practices (and conventions regarding the ordering of authors) differ so widely across disciplines that citation counts are particularly poor methods of allocation across areas.

Peer Review

Ultimately, we believe, efficient QR funding allocation requires an element of peer review. Peer review has its defects; for the same reasons as editorial decisions, for example, it is known to be error-prone when applied to a single piece of work. Across institutions and within disciplines, however, errors should compensate. Peer review is intrinsically more current than other methodologies; unlike the others, it has the chance to be forward-looking. We would also favour a significant international element in any peer review process. This would enable the more objective kind of cross-referencing over disciplines that is rarely seen (for an interesting counter-example, however, see Bessant et al, 2003). Finally, we see merit in peer review being exercised more frequently and less intrusively.

Annexe: Measuring Efficiency in the Water and Sewerage Industry

Prepared by Michael Waterson

Date of Draft 20 June 2006

Summary. The water industry has attracted much investigative attention, and alternative efficiency metrics have been produced, but these seem to lack robustness. The example is relevant since volume and quality are more obviously measurable in the case of water than for academic research.

The methodology of linking outputs to inputs is well established in economics in relation to particular cases, for example the estimation of cost functions, and a policy-relevant example is provided by cost estimation in the water supply industry. The regulator, OFWAT, engages in these calculations in order to reward efficiency – to act as a proxy for a market mechanism in an activity that does not lend itself to direct competition through the market.

However, examining this regime also reveals certain lessons. First, the set of cost drivers is not self-evident, and some experimentation and refinement is almost certainly required. Second, there is a tendency to depart from the model to handle special issues, for example, alleged higher costs in London, in a way that is inevitably somewhat ad hoc. Third, there is likely to be significant volatility in the cross-sectional and time series pattern of the *residuals* in an econometric exercise such as this; residuals are of key importance because they determine who is, and who is not, efficient, given a statistical model. In other words, the process is not very robust.

To illustrate this last point, I tabulate comparisons of the ten water and sewage authorities in respect of the operating efficiency of their water services estimated by the regulator, OFWAT, for 1997/98, and by an academic investigator, Ashton (2000), using a panel from 1987 to 1997. OFWAT puts the ten companies into five efficiency levels A to E (by a process that is not reported in full) and I have done the same for Ashton's results, using the same proportion in each category as in the OFWAT case. I make no claims about the quality of either methodology, but they are designed with the same question in mind.

Company	Ranking:	
	OFWAT	Ashton
Southern	A	C
Wessex	A	B
Thames	B	C
Yorkshire	B	C
Anglian	C	C
North West	C	E
Northumbrian	C	B
Severn Trent	C	C
Dwr Cymru	E	A

It will be apparent that, despite the same aim, the results differ markedly, leading one both to question the robustness of the OFWAT estimates and to be concerned about their use for purpose of calculating comparative rewards.

They continue to be used because a better system has not been devised and because refinements have improved the system. Nevertheless, one might reiterate the point that measuring output in the water industry is much more straightforward than in the higher education research industry.

References

- Ashton, John K. 2000. "Cost Efficiency in the UK Water and Sewerage Industry." *Applied Economics Letters*, 7:7, pp. 455-458.
- Bessant, John, et al. (14 authors). 2003. "The State of the Field in UK Management Research: Reflections of the Research Assessment Exercise (RAE) Panel." *British Journal of Management*, 14:1, pp. 51-68.
- Oswald, Andrew. 2006. "An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-makers." University of Warwick, Department of Economics.