ALDO GEUNA and BEN R. MARTIN

UNIVERSITY RESEARCH EVALUATION AND FUNDING: AN INTERNATIONAL COMPARISON

ABSTRACT. Many countries have introduced evaluations of university research, reflecting global demands for greater accountability. This paper compares methods of evaluation used across twelve countries in Europe and the Asia-Pacific region. On the basis of this comparison, and focusing in particular on Britain, we examine the advantages and disadvantages of performance-based funding in comparison with other approaches to funding. Our analysis suggests that, while initial benefits may outweigh the costs, over time such a system seems to produce diminishing returns. This raises important questions about its continued use.

INTRODUCTION

Research evaluation has emerged as a key issue in many industrialized countries, where universities are faced with demands for greater accountability and the consequences of diminished funding.[1] Universities today are expected to be both efficient and accountable.[2] These pressures have made evaluation essential. In itself, this is nothing new. For more than two decades, there has been growing concern 'about the increasing cost of funding university-based research ... and the need to obtain "value for money" for public expenditure on higher education'.[3] In response, many governments have implemented mechanisms that attempt to relate funding to performance. In this paper, we outline the leading characteristics of university research assessment and funding practices in ten European

[1] OECD, *The Evaluation of Scientific Research: Selected Experiences* (Paris: OECD, 1997).

[2] W.F. Massy (ed.), *Resource Allocation in Higher Education* (Ann Arbor: University of Michigan Press, 1996).

[3] OECD, *Universities Under Scrutiny* (Paris: OECD, 1987). See also C. Gellert (ed.), *Higher Education in Europe* (London and Philadelphia: Jessica Kingsley, 1993); A. Geuna, *The Economics of Knowledge Production: Funding and the Structure of University Research* (Cheltenham: Edward Elgar, 1999); R.G. Noll (ed.), *Challenges to Research Universities* (Washington, DC: Brookings Institution Press, 1998); and OECD, *University Research in Transition* (Paris: OECD, 1998).

Minerva **41:** 277–304, 2003.

countries, Australia, and Hong Kong.[4] We consider some of the advantages and disadvantages inherent in performance-based funding systems; and in the context of Britain, we pose the question, do the benefits outweigh the results over the longer term?

RESEARCH EVALUATION

Although some have attempted to distinguish between 'evaluation' and 'assessment', both terms are used in measuring the qualitative and quantitative outputs of any given academic unit.[5] In practice, 'evaluation' can be divided into *ex ante* and *ex post* forms, and can perform either a summative or formative function.[6] *Ex ante* evaluation is conducted prior to research – to assess its potential significance and likelihood of success. *Ex post* evaluation comes once research has been completed, and assesses output and impact. Summative evaluation involves making judgements about the performance of a unit by comparison with similar units. Evaluation results are increasingly used as inputs in research management.[7] 'Evaluation for strategy' is conducted at both national and institutional levels – in 'quality assessment systems', for example. Evaluation is also used to decide funding, following performance assessments of researchers, projects, programmes, departments, and institutions. The assumption is that funds that are allocated after performance is evaluated, will yield greater returns.[8] In formative evaluation, the aim is to assist a unit in achieving those returns.

There has been much debate about the advantages of evaluation as a tool of research policy.[9] Evaluation in some form inevitably takes place

[4] Information pertaining to these twelve countries covers the period from the mid 1980s to 2000.

[5] P.V. Hills and A.J. Dale, 'Research and Technology Evaluation in the United Kingdom', *Research Evaluation*, 5 (1), (1995), 35–44.

[6] M. Kogan, 'The Evaluation of Higher Education: An Introductory Note', in M. Kogan (ed.), *Evaluating Higher Education* (London: Jessica Kingsley Publishers, 1989), 11–25; Massy, *op. cit.* note 2; and L.E. Suter, 'United States: The Experience of the NSF's Education and Human Resources Directorate', in OECD, *op. cit.* note 1.

[7] J. van Steen and M. Eijffinger, 'Evaluation Practices of Scientific Research in The Netherlands', *Research Evaluation*, 7 (2), (1998), 113–122.

[8] In a few cases, a political decision may be taken to invest in building up weaker groups rather than in concentrating resources on successful ones.

[9] See C. Cooper and D. Otley, 'The 1996 Research Assessment Exercise for Business and Management', *British Journal of Management*, 9 (1998), 73–89; E. El-Khawasi and W.F. Massy, 'Britain's "Performance-Based" System', in Massy *op. cit.* note 2, 223–242;

every time a paper is submitted for publication, or a new professor is appointed or promoted, or a learned society or government body allocates a grant. But while there is a large literature on performance indicators, there is little consensus as to which measurements work best.[10] At the same time, the goals of evaluation tend to be defined by the evaluating agency.[11] In the UK, this is the responsibility of the Higher Education Funding Councils (HEFCs),[12] while in The Netherlands, evaluations are carried out by the Association of Netherlands Universities (VSNU). The HEFCs use evaluation as a method of allocating funds, while VSNU uses evaluation as a management tool. Different agencies also employ different criteria. They tend to focus on four typical output measures: volume, quality, impact, and utility. Peer review and bibliometric measures are their main methods. In 'peer review', the unit of assessment is normally the 'project' or the 'individual'. However, because bibliometric analyses cannot be usefully applied across the board, to all departments in a large number of universities,[13] peer review has become the principal method of university assessment as well.[14] When supplemented with publication and citation data and other information, this method is called 'informed peer review'.

A. Geuna, 'The Changing Rationale for European University Research Funding: Are There Negative Unintended Consequences?' *Journal of Economic Issues*, 35 (3), (2001), 607–632; M. Kogan, 'The Treatment of Research', *Higher Education Quarterly*, 52 (1), (1998), 48–63; S. Kushner, 'The Research Assessment Exercise versus Development in Higher Education: A Response to Richard Pring', *British Journal of Educational Studies*, 44 (1), (1996), 5–8; and G. Whittington, 'The 1996 Research Assessment Exercise', *British Accounting Review*, 29 (1997), 181–197.

[10] Geuna, *op. cit.* note 3. For an early study of difficulties in constructing research performance indicators, see B.R. Martin and J. Irvine, 'Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy', *Research Policy*, 12 (2), (1983), 61–90. For further discussion, see M. Cave, S. Hanney, M. Henkel, and M. Kogan, *The Use of Performance Indicators in Higher Education*, Higher Education Policy Series (London: Jessica Kingsley Publishers, 1997), ch. 4; and R.T.H. van Raan (ed.), *Handbook of Quantitative Studies of Science and Technology* (Amsterdam: North Holland, 1988).

[11] Steen and Eijffinger, *op. cit.* note 7.

[12] The HEFCs are responsible for allocating teaching and general research funds to universities.

[13] The main problem consists in having to 'clean up' institutional addresses, a task that can take many person-years of effort. See B.R. Martin and J.E.F. Skea, 'Academic Research Performance Indicators: An Assessment of the Possibilities' (Brighton: Science Policy Research Unit, 1992).

[14] Surveys suggest that researchers favour peer review over other assessment methods (see *ibid.*). See also S. Cole, J.R. Cole, and G.A Simon, 'Chance and Consensus in Peer Review', *Science*, 214 (1981), 881–886.

Let us consider some of the similar and contrasting approaches to evaluation – first in Europe, then in Asia-Pacific.

University Research Evaluation in Europe

The United Kingdom

In the UK, university research is financed through a dual-support system, which combines general institutional funding with grants and contracts. During the late 1980s, the system underwent radical change. The turning point came with the Education Reform Act of 1988, which created two new agencies – the Universities Funding Council (UFC) and the Polytechnics and Colleges Funding Council (PCFC). With these came a new commercial 'logic', by which the two agencies became 'buyers' of academic services. The idea was that universities, polytechnics, and colleges would be transformed from public institutions, run at state expense, to suppliers, servicing 'clients'. In the early 1990s, the UFC and PCFC were merged into a single Higher Education Funding Council (HEFC); and in 1993, separate agencies were created for England, Scotland, Wales, and Northern Ireland. In 1992, following the publication of an influential White Paper, entitled, *Higher Education: A New Framework*, Britain's polytechnics were granted university status, and the Conservative government began to encourage competition between the 'old' universities and the former polytechnics.

Over the past decade, the UK has developed one of the most advanced research evaluation systems in Europe.[15] Evaluation now takes place not only at the level of the individual researcher and project, but also at institutional and national levels.[16] The first Research Assessment Exercise (RAE) was carried out in 1986, and a similar exercise was repeated in 1989, 1992, 1996, and 2001. In 2001, the RAE was carried out jointly by the four higher education funding bodies.[17] Their aim was to give each unit of

[15] Hills and Dale, *op. cit*. note 5.

[16] As in most industrialized countries, Research Councils and other funding agencies conduct peer review evaluations of individual proposals for new research projects and (in many cases) of completed research projects.

[17] The four are the Higher Education Funding Council for England (HEFCE), the Scottish Higher Education Funding Council (SHEFC), the Higher Education Funding Council for Wales (HEFCW), and the Department of Education for Northern Ireland (DENI). The acronym HEFC is used here to denote all four.

university research[18] a quality rating,[19] on which the distribution of HEFC funds would be based.[20] The RAE's definition of 'research' was broad, and included

> original investigation undertaken in order to gain knowledge and understanding. It includes work of direct relevance to the needs of commerce and industry, as well as to the public and voluntary sectors; scholarship;[21] the invention and generation of ideas, images, performances and artefacts including design, where these lead to new or substantially improved insights; and the use of existing knowledge in experimental development to produce new or substantially improved materials, devices, products and processes, including design and construction. It excludes routine testing and analysis of materials, components and processes, e.g. for the maintenance of national standards, as distinct from the development of new analytical techniques.[22]

The RAE has so far made no separate assessments of basic and applied research. To refute the criticism that it is biased against applied research,[23] panels have been instructed to give equal weighting to all research, whether basic or applied;[24] and to focus upon quality. In response to criticisms that interdisciplinary research has not been fairly assessed,[25] universities have been encouraged to submit interdisciplinary work to the most appropriate panel, and to suggest second panels to consider submissions in parallel.

How can these procedures be characterized? The RAE can be described as an '*ex post* evaluation' based on 'informed peer review'. All research activities within a university are categorized into so-called 'units of assessment' (UoA). In 2001, sixty-eight UoAs were defined, broadly similar to those classified in 1992 and 1996. For each UoA, a panel of ten to

[18] The RAE excludes teaching. Another exercise, the 'Teaching Quality Assessment (TQA)', assesses university teaching.

[19] All university departments are eligible to participate in the RAE.

[20] In the fiscal year 1999–2000, 97% of the £855 million of HEFCE research funds was distributed according to the RAE results. See J. Enderby, 'Excellence Comes Through Diversity', *The Times Higher Education Supplement*, 1472, 2 February 2001, 20.

[21] 'Scholarship' is defined by HEFCE as the creation, development, and maintenance of the intellectual infrastructure of subjects and disciplines.

[22] HEFCE, *Research Assessment Exercise 2001: Guidance on Submissions, RAE 2/99* (Bristol: Higher Education Funding Council for England, 1999).

[23] E.g., J. Griffith, *Research Assessment: As Strange a Maze as E'er Men Trod* (London: Council for Academic Freedom & Academic Standards, Report No. 4, 1995).

[24] HEFCE, *Research Assessment Exercise 2001: Assessment Panels' Criteria and Working Methods, RAE5/99* (Bristol: Higher Education Funding Council for England, 1999).

[25] Royal Society of Chemistry, *Chemistry in the UK – Will it Survive? Conclusions of the Royal Society of Chemistry Workshop* (London: Royal Society of Chemistry, 1995); N. Loder, 'Funding Penalty For Cross-Boundary Work', *Nature*, 399 (1999), 94; and J. Tait, 'Help for the Academic Nomads in Search of Their Own Sympathetic Tribe', *The Times Higher Education Supplement*, 1374, 5 March 1999, 34–35.

fifteen experts was chosen. Some 1,300 professional associations and learned societies were asked to nominate candidates for these panels; panel chairs were nominated by panellists from earlier exercises, and appointed by the Funding Councils. Chairs in turn chose members from nominees proposed by outside bodies, taking into account experience, standing, and representation of user communities.[26]

In the RAE, every department or group within a university is assigned to a UoA, hence to a panel.[27] Information on performance requested in 2001 included the following:[28]

> an overall staff summary – with information on all academic and support staff, whether or not described as 'research active' staff;[29]
> details on research-active staff whose work was to be evaluated;
> publications and other public outputs – for each research-active member of staff, up to four items could be submitted;[30]
> an overview of research students and research studentships;
> details of external research income, including amounts and sources;
> a description of the research environment, its structure, policies and strategies; and
> general observations and additional information (including indicators of excellence).

Using this information, panels judged the quality of each department and assigned a rating on a scale from 1 to 5*. This was used by the HEFCs to determine funding for each unit, with the total block grant calculated by summing across all units.[31]

In 2001, as earlier, publications constituted the core of university assessment. The 1992 RAE required statistics on publications from UoAs, in addition to four published/public outputs from each researcher.

[26] HEFCE, *Research Assessment Exercise 2001: Membership of Assessment Panels, RAE 3/99* (Bristol: Higher Education Funding Council for England, 1999).

[27] Occasionally, a university department may be assigned to more than one UoA, or two departments may be combined in a single UoA.

[28] HEFCE, *op. cit.* note 22.

[29] A department is not obliged to submit the research outputs of all its staff; it can decide to submit only certain 'research active' staff, the implication being that it will receive funds only for those researchers.

[30] The work was to have been published during the period 1 January 1994 to 31 December 2000, in the case of the arts and humanities, and between 1 January 1996 and 31 December 2000 for all other subjects.

[31] In the 2001 RAE, there was a substantial increase in the number of departments awarded 5 and 5* rankings. However, because there was no proportional increase in government funding, the resources allocated to all departments, except those receiving a 5* rating, were cut substantially.

However, the 1996 and 2001 RAEs required only up to four outputs per active staff member. The change was intended to focus upon quality rather than quantity. It also reduced the incidence of 'rush publications' resulting from last-minute attempts to increase aggregate totals.[32]

The Netherlands

In The Netherlands, university research is also financed through a dual-support system. The Dutch Ministry of Education and Science provides core funding through a so-called 'first-flow'. 'Second-flow' grants come from research councils and foundations, and 'third flow' contracts come from government departments and other organizations. As elsewhere, concern with quality and relevance has led to measures of accountability. In 1979, a White Paper recommended changes in the management of research, and in 1983, a system of 'conditional funding' was introduced, ostensibly to make research more efficient and socially relevant.[33] By this scheme, a distinction was drawn between funds given for teaching and research (termed, respectively, 'A-part' and 'B-part' funds). Research positions were financed according to quality of output. Universities were required to produce medium-term research plans, identifying areas of strength that deserved priority. In addition, national objectives were identified, with the intention of increasing research in key fields. In 1992, thirteen universities and the Minister for Education agreed that the Association of The Netherlands Universities (VSNU) should develop a system 'Quality Assessment of Research', using external evaluation to complement internal quality controls.[34] This replaced the system of conditional funding.[35]

In 1993, the 'conditional funding' scheme was replaced by the so-called HOBEK (*Hoger Onderwijs BEKostigingsmodel*, or 'higher education funding model'). This weighted allocations in terms of teaching

[32] There is some evidence that the use of total numbers of publications as a performance measure may have led to 'publishing inflation' – i.e., maximizing the number of articles produced by repetition, lowering quality standards, or the 'salami slicing' of research into 'least publishable units'. See Cave *et al*., *op. cit.* note 10.

[33] Minister for Education and Science, *Beleidsnota Universitair Onderwijs* (The Hague: Minister for Education and Science, Minister for Science Policy and Minister for Agriculture and Fisheries, 1979). See also J. Irvine, B. Martin, and P.A. Isard, *Investing in the Future: An International Comparison of Government Funding of Academic and Related Research* (Aldershot: Edward Elgar, 1990).

[34] The Dutch university system consists of fourteen universities – nine general, three technical, one agricultural, and an Open University (which, because of its different nature, is not included in the scheme described here).

[35] Steen and Eijffinger, *op. cit.* note 7.

(23%), research (64%), and what was called 'inter-weavement' (13%). Funding was based on four factors: student numbers, numbers of previous year degrees, numbers of students completing the degree in the required four years, and numbers of graduating research students. However, the budget was allocated incrementally, and on an historical basis, rather than according to quality.[36] In 1999, HOBEK was replaced by a new model called STABEK ('*STAbiele BEKostiging*', or 'stable funding'). Under this scheme, the government approves funding for several years, so as to ensure greater 'stability'. However, this is intended to be temporary, until such time as the Ministry of Education, Culture and Science develops a scheme that puts greater emphasis upon performance.[37]

In The Netherlands, university research is classified according to disciplines and programmes.[38] The Dutch have used evaluations not to allocate funds, but to develop strategies. On the one hand, as Arie Rip and Barend van der Meulen have shown, the Dutch research culture prefers informal, 'bottom-up' assessments. On the other, policy-makers are more interested in making strategic choices than in evaluating performance.[39] Thus, each of twenty-seven disciplines is evaluated by a different committee. Unlike British practice, according to which all disciplines are evaluated simultaneously, the Dutch phase their evaluations over four to six years.[40] In 1993, a first pilot group of disciplines was chosen, each with a Review Committee of five to seven experts, set up by VSNU in consultation with the Royal Academy of Arts and Sciences (KNAW). The chair was either a Dutch expert or a foreign expert thoroughly familiar with the Dutch scene. To ensure impartiality, committee members were predominantly foreign. For this reason, the principal language was English, and the results were published in English.

These committees continue to evaluate performance over periods of five years, in the following categories:

---

[36] J.B.J. Koelman, 'The Funding of Universities in The Netherlands: Developments and Trends', *Higher Education*, 35 (2), (1998), 127–141.

[37] P. van der Meer, 'Funding and Allocation of Resources in Higher Education: The Dutch Case', paper prepared for the 3rd ALFA-BRACARA international conference on 'Funding and Allocation of Resources in Higher Education', held on 8–10 February at Universidad Nacional Autonoma de México (1999).

[38] These represent the smallest unit of assessment – hierarchically, research programmes are organized under university departments.

[39] A. Rip and B.J.R. van der Meulen, 'The Patchwork of the Dutch Evaluation System', *Research Evaluation*, 5 (1), (1995), 45–53.

[40] Because evaluation is not used for determining the allocation of funds, there is no need to evaluate all disciplines at the same time. The first round of evaluations covering twenty-seven disciplines was completed in 1998, and a second round began in 1999.

academic staff;
programme mission and research plan;
content of programmes and main results;
publications;
five selected key publications; and
other indicators of quality and reputation (such as patents and invited lectures).

The committees also conduct site visits and interview programme leaders. Where possible, they commission bibliometric analyses,[41] because VSNU believes that these complement written and oral information.[42] The committees then assess each programme in four dimensions:

1. scientific quality – originality of ideas and methods, importance of the discipline, impact, and prominence;
2. scientific productivity – inputs and outputs (staff and funds are inputs; while outputs are number (and nature) of publications, dissertations, patents, and invited lectures);
3. scientific relevance – relevance to the advancement of knowledge and technology; and social consequences; and
4. long-term viability for research, publication, coherence, and continuity of research.

The assessment is translated into a five-point rating (1 = 'poor', 5 = 'excellent'). In the most recent 1999–2002 assessment, the framework remained the same, but greater leeway was given to assess groups in relation to their differing missions. (Quality and productivity continue to be assessed according to a single standard, as before.) This greater flexibility reflects VSNU's recognition of differences between institutions. However, it remains to be seen how much effect this distinction will have, since the assessment committees are free to decide the extent to which they take differing missions into account.[43]

Germany

In Germany, most academic research is conducted either in organized research institutes (such as the Max-Plank or Fraunhofer institutes) or

[41] For example, the Review Committee for Chemical Research was provided with a bibliometric analysis produced by the Centre for Science and Technology Studies at Leiden University (CWTS). This evaluated the impact of journal articles from all the research programmes assessed by the committee.

[42] VSNU, *Quality Assessment of Research: Chemistry, Past Performance and Future Perspectives* (Utrecht: VSNU (Association of Netherlands Universities), 1996).

[43] Personal communication from Anne Klemperer (June 1999).

in the 'scientific universities' (Wissenschaftliche Hochschulen). Some research is also carried out in the polytechnics (or Fachhochschulen), which are, however, mainly teaching institutions. Their research is funded by contracts from industry or government agencies. There are only a few private universities (such as the Universtät Witten-Herdecke).

There are three categories of public funding for university research. The first is institutional funding, which takes the form of block grants from the state (Bundesland); this constitutes almost two-thirds of total university expenditure, and covers basic infrastructure and staff. The second comes in the form of a capital grant for buildings and large-scale equipment, and is provided jointly by the federal government (Bund) and the Länder. The third source is 'third party funds' (Drittmittel), which are grants and contracts given by public institutions for specific projects. A large proportion of these are allocated by the Deutsche Forschungsgemenschaft (DFG), which is financed jointly by the Bund and the Länder. Smaller proportions are given by industrial sponsors, such as the Volkswagen-Stiftung and the Fritz-Thyssen-Stiftung. As in other countries, such funds are granted on the basis of peer review, using criteria of scientific excellence and social relevance.

Institutional and capital funds are allocated according to a profile that includes numbers of students and staff and current spending. To determine research budgets, an 'R&D coefficient' is derived from surveys, showing time spent on research and teaching. In general, performance measures have not been used to allocate research funds, and there have not been evaluations for this purpose.[44] In recent years, however, a few Länder have allocated additional resources on a competitive or performance-related basis. In particular, Lower Saxony set up a commission in 1998 to assess performance in the Länd's twelve universities.

There has as yet been no federal evaluation scheme covering all German universities. This is partly explained by the fact that the universities are financed mainly by the Länder, but it also reflects widespread resistance to the idea of inter-university competition. Among German academics, 'competition is not yet seen as a principle for advancing and encouraging research (and teaching) quality',[45] although some Länder have started inter-university competitions in certain faculties (e.g., in the medical faculties of Baden-Württemberg). Consequently, although there have been

[44] Research evaluations of research institutes have become more common since 1990.

[45] D.F.J. Campbell and B. Felderer, *Evaluating Academic Research in Germany. Patterns and Policies*, Political Sciences Series No. 48 (Wien: Institute for Advanced Studies, 1997).

some evaluations of university research, these have mostly not influenced funding.[46]

In the late 1990s, the Federal government began to press for evaluation.[47] However, the German constitution grants universities considerable autonomy. Indeed, it has been argued that, with the exception of student evaluations, systematic governmental evaluations of university professors are unconstitutional. Nevertheless, individual universities (and individual faculties) have begun evaluations of their own. The Freie Universität in Berlin, for instance, has done so, and has used its results in internal funding distributions.[48] However, other universities – such as those in the Verbund Norddeutscher Universitäten (VNU) – see this as a merely retrospective gesture, in that it rewards those that already have good records, while what is needed is a mechanism that helps universities to improve their performance.

In 1998, German higher education underwent a major reform with the Bundestag's adoption of an amendment to the Framework Act, which makes competition possible through deregulation, performance orientation, and the creation of incentives.[49] This provides a legal basis for important structural improvements, including the introduction of financing based on teaching and research. It abolishes the previous 'immunity' of professors to external evaluation. Hence, external evaluation may now begin to develop. Furthermore, some Länder (such as Baden-Württemberg) are rethinking the traditional idea of the all-encompassing university with a complete set of faculties and institutes. Instead, the future may see new 'centres of competence' only at certain universities in any given Länd.

Scandinavia

The Nordic countries use a dual-support system to finance university research. In Denmark and Finland, research councils give grants based upon international peer review, and governments give institutional funds. Only in Denmark and Finland do institutional funds contain a performance-based component. In Denmark, a new budgeting system was introduced in 1994, in which a distinction is made between funds for teaching and for research. The budget contains five elements: a basic grant,

46 H.D. Daniel and R. Fisch, 'Research Performance Evaluation in the German University Sector', *Scientometrics*, 19 (5–6), (1990), 349–361.

47 Campbell and Felderer, *op. cit.* note 45.

48 *Ibid.*

49 BMBF, *Facts and Figures 1998* (Bonn: Bundesministerium für Bildung und Forschung, 1998).

a performance-related grant for teaching, a research grant, a grant for other activities, and a capital grant. Until 1995, research funds were allocated on an incremental basis. Since then, amounts awarded have depended upon the volume of teaching and external research income. No other performance measures are used, although PhD student numbers help determine the performance-based grant for teaching.

In Norway, universities receive block grants, and no distinction is made between teaching and research. Until recently, there has been no attempt to adopt performance measures, apart from giving a fixed sum per doctoral graduate. Similarly, in Sweden there is no mechanism for performance-based research funding. Following legislation in 1993, examination results are the only indicator considered in deciding the allocation of funds. Among the Scandinavian countries, Finland has had the most experience of performance-based funding,[50] so let us consider its development.[51]

Most academic research in Finland is conducted in the Finnish universities, and is financed by core funding from the Ministry of Education; by peer-reviewed grants awarded by the four research councils under the Academy of Finland; and by research contracts from industry and government.[52] The Ministry of Education uses a system known as 'Management by Results', introduced in 1994, in which a small proportion of every budget is based on an assessment of performance. This was followed by the adoption in 1998 of three-year agreements that specify the outcomes that each university is expected to achieve, and the levels of funding that each receives. These three-year agreements are updated annually.

At present, the agreed areas of expenditure comprise basic (90%), project (7%) and performance-related (only 3% at present, but expected to increase funding). Basic funding covers salaries and facilities. A formulaic model for basic funding has been used since 1997, in which a connection between teaching and research is made explicit. The teaching component is represented by target numbers of Master's degrees, and the research element by target numbers of doctoral degrees. Project funding is earmarked for programmes that the government defines. Performance-related funding is awarded on the basis of quality and impact indicators, and has been used to establish centres of excellence;[53] to increase interna-

[50] E. Helander, 'Evaluation Activities in the Nordic Countries', *Scientometrics*, 34 (3), (1995), 391–400; and T. Luukkonen, 'The Impacts of Research Field Evaluations on Research Practice', *Research Policy*, 24 (3), (1995), 349–365.

[51] This section is based on information derived from Ministry of Education, *Management by Result, http://www.minedu.fi/eopm/hep* (no date, accessed 15 November 2001).

[52] Polytechnics neither engage in research nor offer postgraduate education.

[53] In 1993, a policy to create centres of research excellence was introduced. A 'centre of excellence' could be either a research group or centre, or a larger umbrella organization or network. The Academy of Finland, FINHEEC, and the Ministry of Education

tional collaboration; to improve graduate placement; and to meet planning targets.

The Academy of Finland has been evaluating research performance since the 1970s, focusing mainly on individual scientists, projects, and teams. Initially, there was little by way of a systematic, nationwide *ex post* evaluation of research,[54] However, beginning in the 1980s, there have been evaluations of about twenty fields, including inorganic chemistry (in 1983), automation technology (1986), legal science (1994), and molecular biology and biotechnology (1997). These have focused on international outcomes. They have been driven more by a desire to improve the quality of science than by the need to make funding decisions.

The Finland Higher Education Evaluation Council (FINHEEC), established in 1995, conducts evaluations of three main kinds: institutional; programme/thematic; and accreditation. None of these is targeted specifically at research, however. Nor do institutional evaluations use a uniform model for all universities. On the contrary, the government recognizes differences between universities and emphasizes the developmental role of evaluation. As a result, most evaluations are broad assessments of basic preconditions for teaching and research and the capacity for change. These include statements of an institution's mission, processes, institutional arrangements, resources, and performance. Less attention is paid to the latter two factors. Emphasis varies across universities; one might highlight its teaching, another, its regional role.

FINHEEC evaluations take place in three phases. First, each university carries out a self-evaluation and prepares a report, which is assessed by an external team that visits the university and then produces a final report. Academy evaluations, however, proceed differently. In the case of electronics research, for example, the Academy of Finland commissioned an evaluation in 1985 from the Research Council of Natural Science and Engineering. A committee was set up and two international experts appointed. The scope was limited to certain pre-defined sub-areas, in which twenty-eight university groups were identified and evaluated with respect to:

---

are responsible for the selection, which involves international evaluation in six categories: national/international standing of researchers; scientific significance (i.e., innovativeness and effectiveness of research); quality, quantity, and focus of scientific production; patents; national and international mobility of researchers; and number and level of foreign researchers. By 1997, seventeen research units had been designated centres of excellence and received extra funds. Recently, the differing nature of disciplines has been taken into account with general criteria being adjusted to each.

54 B. Felderer and D.F.J. Campbell, *Evaluation der Akademischen Forschung* (Vienna: Institute for Advanced Studies, 1998).

1. mission, vision, and goals;
2. efficiency in using resources;
3. scientific competence and degree of innovativeness;
4. technological competence and cooperation with other researchers, industry, and users;
5. the national/international importance of results for the scientific community; and
6. the relevance of a group's research for industry.

As with FINHEEC, evaluations were conducted in three phases. A questionnaire was distributed to groups and, having examined the results, evaluators interviewed each group, summarizing their findings (in English). Groups were given opportunity to comment. In a report entitled, 'Management by Result', the Ministry of Education proposed a performance-based mechanism similar to that of the UK's RAE, and suggested that 35 per cent of funds should be allocated on the basis of research performance.[55] It advocated that all university groups be evaluated by the Academy of Finland every three years, using peer review, with research units being graded on a five-point scale, which should be used to determine the funds they receive. The suggestion was criticized by almost all the universities, and the proposal was 'frozen' by the Ministry. The main objection was that the mechanism would give the Academy undue influence.[56]

Eastern Europe: Hungary, Poland, and the Slovak Republic

After 1945, the national research systems in Eastern Europe were based on the Soviet tripartite model, according to which the universities focused on teaching; basic research was conducted in Academy institutes; and applied research was done either in Academy institutes or government ministries.[57] Governments also adopted the Soviet system of block grants to institutes, by which scientist-administrators had great power over the distribution of funds. Favouritism often led to poor research.[58] During the last ten years, with the development of open market economies, this system has changed. The autonomy of science, recognized by self-evaluation and peer review – but subordinated to central planning during the Communist era – has been restored. Research evaluation has emerged as a tool to examine how and

[55] E. Kaukonen, 'Evaluation of Scientific Research in Finland', in OECD (ed.), *The Evaluation of Scientific Research: Selected Experiences* (Paris: OECD, 1997), 12–25.

[56] *Ibid.*

[57] M.S. Frankel and J. Cave, 'Introduction', in M.S. Frankel and J. Cave (eds.), *Evaluating Science and Scientists: An East-West Dialogue on Research Evaluation in Post-Communist Europe* (Budapest: Central European University Press, 1997), 1–6.

[58] *Ibid.*

where to cut budgets without destroying research. Peer review has become the main instrument.[59]

Whilst national systems have experienced major changes, the Academies continue to play a major role in research, and evaluations have focused on the institutes.[60] Almost every year, for example, the Hungarian Academy of Sciences has conducted a comprehensive review of its institutes. The evaluation in 1992 had a special importance because it coincided with an economic crisis and a cut-back in funds. The study evaluated each of the several institutes, using peer review and quantitative indicators. The idea was to support a more selective distribution. This led to a number of recommendations concerning the Academy's network, its management of resources, and the need for organizational change.[61]

In the Slovak Republic, the Academy of Sciences and the universities set up an accreditation committee in 1992 to evaluate Academy institutes and departments. For the institutes, the following indicators were used:

1. publications during the previous five years, classified by type, with ten representative publications;
2. citations (in the *SCI*) during the previous five years;
3. memberships of editorial boards; and
4. participation in conferences, membership of international science organizations, and relations with the international science community (e.g., co-operative agreements and joint projects).[62]

Based on these data, the institutes were graded. In the case of university departments, the same indicators were supplemented by:

1. qualifications of teaching staff;
2. numbers of postgraduate students;
3. research activities of senior staff; and
4. subsequent employment of graduates, together with numbers of applications from foreign students.

Since 1992, such evaluations have been conducted every three years. Faculties are classified into one of four grades which, when combined with student numbers, determine the amount of money each receives.[63]

[59] See K. Hangos, 'The Limits of Peer Review: The Case of Hungary', in Frankel and Cave (eds.), *op. cit.* note 57, 71–81; and P. Zilahy and I. Láng, 'The Evaluation of Research Institutes in Hungary', in Frankel and Cave (eds.), *op. cit.* note 57, 82–92.

[60] UNESCO-CEPES, *Ten Years After and Looking Ahead: A Review of the Transformations of Higher Education in Central and Eastern Europe* (Bucharest: UNESCO-CEPES, 2000).

[61] Zilahy and Láng, *op. cit.* note 59.

[62] J. Tino, 'Institutionalizing Evaluation of Research and Education in the Slovak Republic', in Frankel and Cave (eds.), *op. cit.* note 57, 166–169.

[63] Personal communication from Stefan Zajac (June 1999).

In Poland, legislation in 1991 set up a new system for managing research, led by a Committee for Scientific Research (CSR).[64] The chairman of the CSR is appointed by Parliament, and two-thirds of its members are researchers elected by the scientific community (the remainder being ministers). The CSR is responsible for science policy, and for the distribution of funds through competitive channels. All institutions, including university faculties, compete for funds through the CSR.

Polish universities compete for funds on the basis of student numbers and through two CSR schemes. The first is a grant system for individuals and teams, based on open competition and peer-review. The second is so-called 'statutory funding', which is distributed to faculties within universities on the basis of *ex post* evaluations. Each year, institutions submit their past year's achievements and a research plan for the coming year. Assessments are conducted by expert panels who assign institutions to a category. Allocations are decided by the CSR's Committee for Basic and Applied Research.[65]

Until recently, funding levels were determined by an algorithm using a combination of quantitative and qualitative factors. However, the latter were criticized for their subjectivity, and in 1998 a new formula was introduced. This new 'parametric system' is almost entirely quantitative. It consists of a number of points given for *performance,* R_p, and for so-called *general results,* R_g. The total number, $R = R_p + R_g$, is divided by the number of staff (N) to yield an indicator of effectiveness (E). This is the basis for, every three years, classifying institutions into one of five grades, and for determining their level of funding. Recently, however, the new formula has also been challenged, and it will probably be changed.

Research Evaluation in the Asia-Pacific Region

Much can be learned from recent developments in the Asia-Pacific region. In 2000, New Zealand began to allocate 20 per cent of its institutional core research funding on the basis of peer-review evaluation (with the remainder based on student numbers). If this experiment proves successful,

[64] This discussion is based on J. Jablecka, 'Changes in the Management and Finance of the Research System in Poland: A Survey of the Opinions of Grant Applicants', *Social Studies of Sciences*, 25 (4), (1995), 727–753, and on personal communication with J. Jablecka (June 1999).

[65] The relative proportion of funds received in classes A, B, and C is, respectively, 1.151, 0.951, and 0.550 – see Committee for Scientific Research (KBN), *http://www.kbn.gov.pl/en/general/reseval.html* (accessed 31 August 2001). Institutes of the Polish Academy of Sciences and other government research establishments are financed according to the same formula and evaluation criteria as university faculties.

it is intended to increase the proportion from 20 to 80 per cent. In Australia and Hong Kong, there are national research evaluations, and both use their results in allocating funds.[66] Assessment in Australia is concerned mainly with research grant incomes and research output indicators, but Hong Kong uses a scheme similar to the UK's RAE.

Australia

Since a so-called 'Unified National System' was introduced into Australia in 1988, the number of universities eligible for research funding has nearly trebled. The Commonwealth government provides funds through a dual-support system,[67] consisting of an institutional operating grant and specific research grants. Institutional funds are currently given as block grants by the Department of Education, Science and Training, with universities having constitutional discretion over internal distribution. Grants for specific projects are awarded by agencies such as the Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC). The former operates the National Competitive Grants Program, which includes funding for 'Discovery Projects' and for 'Linkage Projects' aimed at supporting collaborative research with industry partners.

Peer review is the main method for allocating specific grants, but a formula is used to distribute core funding.[68] Under a 'Relative Funding Model', introduced in 1990, funding for teaching was based on student numbers, while research support was measured by a 'Research Quantum', based on success in winning Commonwealth Competitive Grants. When it was recognized that this did not fully represent research performance, the criteria were broadened to include other sources of funding. In addition, measures other than funding – such as publications and higher degree completion rates – were incorporated in the formula.

In 1993, the Minister for Education announced that, as from 1995, the Research Quantum would be allocated on the basis of a new Composite Index. From 1998, the Department of Education, Training and Youth Affairs (DETYA), and subsequently the Department of Education, Science and Training (DEST), has been responsible for collecting data for the

66 H. Atkinson and W.F. Massy, 'Quantitative Funding Models', in Massy, *op. cit.* note 2, 245–266; and P. Bourke, *Evaluating University Research: The British Research Assessment Exercise and Australian Practice*, Commissioned Report No. 56 (Canberra: National Board of Employment, Education and Training, 1997).

67 Over 80% of university research is funded by the Commonwealth government. Other funders are state and local governments, industry, non-profit organizations, and overseas providers – see Industry Commission, *Research and Development Report* No. 44 (Canberra: AGPS, 1995), vol. 1.

68 *Ibid.*

Composite Index; for calculating allocations; and for advising the Minister on the Index and the weighting of elements within it. Both input and output measures are incorporated in the Composite Index, as follows:

1. research inputs (funding):
   a) each university's funding from Commonwealth competitive grants;
   b) other public-sector research funding; and
   c) industry and other research funding.
2. research outputs:
   a) scholarly publications by staff and students; and
   b) higher degrees (Masters and PhDs) completed.

The components of the Index have been weighted differently from year to year. If, for example, university X's shares of national funding, publications, and completed higher degrees, averaged over the last two years, are 4.5 per cent, 3.6 per cent, and 5.3 per cent, respectively, then its Composite Index is obtained by multiplying the share by the corresponding weighting for each component (80%, 10%, and 10% respectively in 1999), and then adding them together. The Composite Index of University X represents its share of the total research activities for all universities, and its Research Quantum allocation is the Composite Index multiplied by the total Research Quantum available.

As compared with the UK's RAE, the Australian Research Quantum system is mechanistic. Research performance is evaluated solely on the basis of quantitative measures – that is, on volume, not quality. It is 'confined to counting annually the gross number of undifferentiated entrants each institution can place in the classification categories which are weighted for funding purposes'.[69] In 1994, the Australian Vice-Chancellors' Committee proposed that qualitative aspects should be incorporated. This was taken further in 1997 by Paul Bourke, who noted that the UK's RAE fulfils three functions – serving as a competitive source of discretionary income, as a reward for quality and/or volume of output, and as an instrument of policy. By contrast, the Research Quantum performs only the second of these; it does not reward quality, and is seldom used as an instrument of policy. Bourke proposed a funding mechanism that would combine the Research Quantum with the 'professional judgement' element of the RAE.[70] However, this has not materialized. It is argued that, as funding generated through the Research Quantum has become vital to the universities, any major change will need careful justification.[71]

69 Bourke, *op. cit.* note 66, 25.
70 *Ibid.*
71 *Ibid.*

Hong Kong

Higher education in Hong Kong is publicly funded through the University Grants Committee (UGC), a non-statutory body that acts as adviser to government. Research is funded through a dual-support system.[72] The UGC provides institutional funding, while the Research Grants Council (RGC), which operates under the aegis of the UGC, allocates project grants based on international peer review.[73] The research component of core funding is allocated on the basis of performance, using (since 1993) a mechanism similar to the UK's RAE.[74] In 1993, UK experts gave advice, and Hong Kong adopted the same name – the Research Assessment Exercise.[75]

Three RAEs have been carried out by the UGC – in 1993, 1996, and 1999. Their aim has been 'to measure the output and quality of research of the UGC-funded institutions by cost centre as the basis for allocating some of the research portion of the institutional recurrent grant for the next triennium in a publicly accountable way'. In 1999, the unit of analysis was a set of 58 'cost centres', and each university department or research unit was required to map onto these. The cost centres were evaluated by a number of panels, and the results were used to determine institutional research funding levels.

ADVANTAGES AND DISADVANTAGES OF PERFORMANCE-BASED FUNDING

So far, we have identified a continuum of approaches towards research funding. At one extreme, we find the pure performance-based, *ex post* model; at the other, the allocation of resources is based on educational size. Few countries have implemented the first approach, but among them, Britain has led the way.[76] In the 1980s, British research policies aimed at accountability and selectivity. In the 1990s, however, came models based on some form of assessment. In each case, the academic community has

[72] This section is based on N.J. French, W.F. Massy, P.K. Ko, H.F.H. Siu, and K. Young, 'Research Assessment in Hong Kong', *Journal of International Education*, 10 (1), (1998); and UGC, *Research Assessment Exercise 1999, Guidance Notes* (Hong Kong: University Grants Committee, 1999).

[73] Research Grants Committee, 'Refutation of Article in 'Next' Magazine of 9 October 1998', *http://www.ug.edu.hk/RGC/documents/E_XNEXT.html* (accessed 20 July 1999).

[74] UGC, *op. cit.* note 72.

[75] French *et al*., *op. cit.* note 72.

[76] Geuna, *op. cit.* note 3.

resisted certain criteria, and has been guarded in its acceptance of any model that links evaluation with funding.

At present, we lack data on inputs and outputs over a sufficiently long period to assess the success of the several different systems. Nonetheless, we can offer some preliminary thoughts on the advantages and shortcomings of the two principal approaches to university research funding (see Table I).

Given the inherent difficulties in evaluating institutions, there will always be reasons for opposing a performance-based approach to funding. Its main virtue lies in the assumption that it is ostensibly meritocratic, rewarding success and improving quality. A performance-based system can, it is said, increase efficiency in the short term. It may also provide greater accountability. It gives a mechanism to link research to policy, a way to shift priorities across fields, and a rational method of moving resources from less well-performing areas to areas where they can be used to greater effect.[77]

While these arguments have their merits, a performance-based system has drawbacks. First, obtaining reliable and comparable information is costly. Assessments based on peer review are especially labour-intensive, when all a nation's universities and their constituent departments have to be judged. Nor do indicator-based approaches offer a shortcut; if conclusions are to be robust, data must be accurate and reliable.[78]

Second, a performance-based funding system, because it encourages competition, may also encourage a shift towards the 'homogenization' of research, discouraging experiments with new approaches, and rewarding 'safe' research, irrespective of its benefits to society.[79] The resulting decrease in diversity may be harmful. Moreover, a system that has publication as a key criterion encourages 'publication inflation'.[80] Some academics will almost certainly respond by 'game playing' without necessarily improving performance.[81]

Third, performance-based funding can widen the gap between research and teaching. If rewards for research are greater than rewards for teaching, academics will focus on the former at the expense of the latter. While the

[77] Assessments also give leading departments a 'marketing' tool to attract top researchers and students.

[78] P. Bourke, and B. Martin, *Evaluating University Research Performance – What Approach? What Unit of Analysis*? (Canberra: ANU and Brighton: SPRU, 1992).

[79] E.g., A. Thomson, 'RAE Faces Axe in DTI Review', *The Times Higher Education Supplement*, 1408, 29 October 1999, 1.

[80] See note 32.

[81] E.g., D. Cannadine, 'Flaws of Supply with No Demand', *The Times Higher Education Supplement*, 1381, 23 April 1999, 18–19.

TABLE I

Advantages and drawbacks of alternative approaches to university research funding

	Advantages	Drawbacks
Performance-based	– 'meritocratic' in that it links resources to performance, rewarding good research – strong incentive to improve individual as well as institutional performance – competition may lead to increased efficiency – ineffective research identified and cut – encourages research to be properly completed and written up for wider dissemination – provides public accountability for government funds invested in research – encourages more explicit/coherent research strategy on part of department or institution – provides mechanism for linking university research to government policy (e.g., to shift priorities) – concentration of resources may enable best departments to compete with world leaders (e.g., in US)	– high cost and labour intensity (whether peer review or indicator-based) for universities and evaluating agencies – may ⇒ 'homogenization' of research and universities – i.e., decrease in diversity and experimentation – may discourage more innovative and risky research – encourages 'publication inflation' (e.g., 'salami publishing') and other 'game playing' (e.g., with indicators) – i.e., 'looking good' rather than necessarily doing better – may encourage traditional 'academic' research at expense of research linked to society's needs – tends to separate research from teaching ⇒ lower priority for teaching – rewards past performance not current or future potential ⇒ reinforces research elite/status quo – may ⇒ over-concentration – may lead to excessive government influence/'interference' in university research

TABLE I

Continued

	Advantages	Drawbacks
Educational size	– low cost to administer – provides departments with 'seed corn' funds to invest in new people/research areas – provides 'space' for long-term research and scholarship – encourages diversity in research – enables academics at any university (not just an elite few) to get involved in research – encourages integration of teaching and research so can exploit synergy between them – protects autonomy of institutions and individuals	– little direct incentive to improve research performance (whether individual or institutional) – may ⇒ stagnation – may give excessive power to officials who distribute core funding within institution – little public accountability for funds (notionally) provided for research – may ⇒ 'ivory tower' research with no social or other relevance – may reinforce public stereotype that some academics are 'lazy' – may be little or no correlation between student numbers and level of research effort by department – distribution of resources/effort may bear little relationship to stated government policy – spreading resources evenly but thinly may ⇒ unable to compete with world-leading institutions

intention is to reward and encourage, assessment inevitably focuses on past rather than current performance, let alone future potential. In consequence, the status quo is reinforced. Institutions that have done well in the past may continue to attract a disproportionate share of resources, perhaps depriving others of the chance to become leaders in their field. Finally, a performance-based system, by making it easier for government to shift priorities, can lead to an excessive level of government influence. For many, autonomy is vital to the health of research and higher education.[82]

On the other hand, funding systems based on educational size have certain advantages. Such systems are simple and cheap to operate, requiring only comparable data on student and staff numbers. The model implies that institutions and departments have 'seed corn' to invest in staff or new (or weaker) areas. If researchers need not focus exclusively on research assessment, they may be inclined to favour more fundamental, longer-term and 'risky' projects. A system based on size also offers opportunities to all institutions, whatever their reputation, adding to the diversity of approaches. Where research funds depend on teaching, the system encourages the integration of teaching and research to the benefit of both. High-quality teaching attracts students, increases the size of the institution, and the flow of research funds. Finally, a system based on size offers institutions and individuals greater autonomy, a feature that many academics find essential.

There are, however, drawbacks to the 'size' alternative. As Table I shows, the disadvantages of a system based on size mirror the advantages of a performance-based system. A system that awards research resources on the basis of teaching gives little incentive to improve research performance. Over time, this can bring stagnation, a situation familiar in the former Soviet Union and Eastern Europe. Moreover, the system can give excessive power to officials, reflecting 'politics' rather than performance. The system also affords less accountability. It may also encourage 'ivory tower' research, and so reinforce stereotypes about 'lazy' academics. Finally, an institution devoted to teaching that receives as much funding for research as does an institution of similar size that elects to devote itself to research, is bound to raise questions of equity as well as efficiency.

With a system based on size, there is a strong chance that the distribution of research resources may bear little relationship to policy. For example, burgeoning numbers of students in media studies would produce a rapid increase in that field's research resources, irrespective of whether media studies is actually deemed a research priority. Conversely, a reduc-

[82] See J. Ziman, *Real Science: What It Is, and What It Means* (Cambridge: Cambridge University Press, 2000).

tion in numbers choosing to study physics could translate into a reduction in research funding, regardless of whether this is in line with government policy. Above all, the allocation of funds by size is likely, in all but the richest of countries, to result in spreading resources too thinly, with the consequence that no institution is able to compete successfully with world-leading institutions.

DO THE BENEFITS OUTWEIGH THE COSTS?

Over the past decade, many governments have pursued a policy of greater selectivity and concentration. In Britain, as elsewhere, an elitist system of higher education has been overtaken by a system of 'mass' higher education, in which the number of universities has increased. Unless there is a similar increase in funding, research resources will be spread too thinly. However, any such increases are unlikely, so performance-based, research-funding systems could be used to concentrate resources – notably, at a small number of universities, which will then stand a better chance of competing with leading institutions abroad.[83]

Over the past decade, there has been a trend towards such performance-based systems. Do the benefits outweigh the costs? Although the lack of data on inputs and outputs makes a proper cost-benefit analysis difficult, we can begin to offer an answer. When assessment is introduced, it usually meets opposition. Many academics see it as a temporary aberration that will disappear once politicians and 'bureaucrats' recognize its defects. Figure 1 indicates how benefits and costs may vary over time. Most benefits will not accrue for some years – hence, the benefit curve is shown as rising gradually from zero, then accelerating. By contrast, the costs of introducing the assessment will be substantial. A procedure must be designed, negotiated, and agreed. Universities must adopt procedures for compiling information. A peer review system must be constructed, and peers must spend time judging performance. Thus, for an initial period, the costs will probably outweigh the benefits.

Over time, however, the benefits of a performance-based system will grow. When carried out a second time, academics will accept that assessment is here to stay. Some will have begun to reap resources from the first assessment, encouraging them to improve their research. Under this model, weaker groups will be identified, and resources will be shifted to more productive groups, thereby enhancing the 'efficiency' of the system. Universities will develop clearer goals and strategies, although a few years

[83] C. Sanders, 'Research-Rich Elite Wins Cash', *The Times Higher Education Supplement*, 1428, 24 March 2000, 6–7.

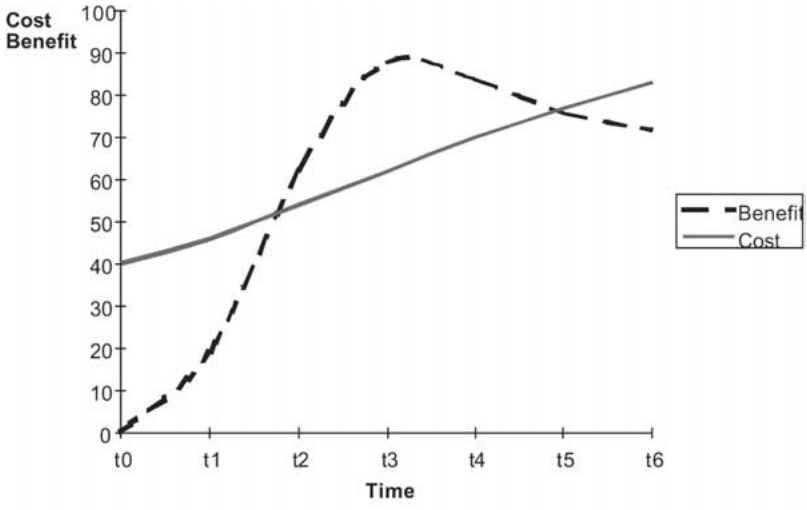

Figure 1.

may elapse before the full benefits emerge. At some point, benefits may come to exceed costs, so the benefit curve in Figure 1 is shown as rising above the cost curve.

Later, however, increases in benefits will begin to level off. Although it is difficult to produce evidence, our impression, based on the UK's experience, is that after a number of exercises, the level of benefits reaches a peak, and then encounters diminishing returns.[84] The conspicuously weak performers have been identified, and the strategies put in place as the first or second assessment have taken effect, so that scope for further gains is limited. Hence, the benefit curve in Figure 1 is shown as falling away, although at an ever-decreasing rate.

[84] This claim might appear to be contradicted by the results of the UK's RAE in 2001, which show a large increase in the percentage of academics in units receiving the top two grades of 5 and 5* (55% compared with 31% in 1996, and 23% in 1992). The Higher Education Funding Councils claim that this represents a genuine improvement in research, citing, in support, the views of the 300 international peers consulted, and unpublished bibliometric data showing that the impact of UK publications has increased since 1996 compared with the world average (Anon., 'Selectivity Raises Scores Across the Board', *The Times Higher Education Supplement*, 1517, 14 December 2001, 3). However, some believe that at least part of the apparent improvement is due to the preparation of better RAE submissions, as academics have learnt to 'play the game' (see Anon., 'Researchers Rise to the Challenge of Excellence', *idem.*, 26). Moreover, other published bibliometric data show no significant improvement in the impact of UK publications over the past 10–15 years compared with the rest of the world. See Observatoire des Sciences et des Techniques, *Science & Technologie Indicateurs 2002* (Paris: Economica, 2002), 203, Table 3.22. Hence, exactly how much of the improved ratings reflects a genuine improvement in performance remains unclear.

What of the costs? Figure 1 shows the cost curve as monotonically (although not necessarily linearly) rising. There are a number of reasons for this. First, as funding comes to depend more upon assessment, methods will be improved, and more effort will be devoted to submissions. Second, as more improve their performance, an 'arms race' or 'Red Queen effect' will result, by which, as competitors become stronger (or learn to 'play the game'), everyone will need to run faster just to stand still. Some of the drawbacks listed in Table I, and their associated 'costs', are likely to discourage risky and longer-term research.[85]

If these assumptions are valid, the benefit curve will at some point fall below the cost curve.[86] From then on, costs will exceed benefits. In the meantime, we have assumed that the peak of the benefit curve rises above the cost curve. But what if this is not the case? Figure 2 suggests what the respective curves might then look like. The benefit curve has a similar shape but a lower peak, so that it never rises above the cost curve. In this case, the cost 'curve' is shown as monotonically decreasing, because in

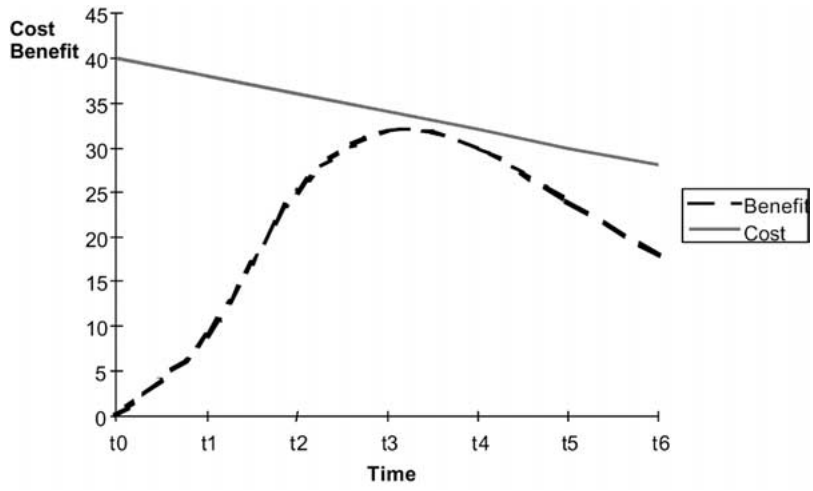

Figure 2.

[85] For example, a survey at Warwick University has revealed that young researchers feel under pressure to choose topics to suit the perceived preferences of RAE panels, and that this pressure was more widespread in 2001 than in the previous RAE. See A. Goddard, 'RAE Under Fire for "Pressurising" Researchers', *The Times Higher Education Supplement*, 1499, 10 August 2001, 4. See also M.J. Larkin, 'Pressure to Publish Stifles Young Talent', *Nature*, 397 (6719), (1999), 467.

[86] Each notional curve is summed over all the institutions participating in the assessment, and takes into account the costs and time of those engaged in carrying out the assessment.

such circumstances, assessment would be seen as not worth the effort. Even if required by funding agencies, academics would merely 'go through the motions'. If such assessments were imposed, universities would find ways of doing less work to satisfy assessors. However, once the benefit curve began to fall off, assessments would be discontinued.[87]

Do the actual cost and benefit curves resemble those in Figure 1 or Figure 2? The short answer is that we cannot be sure. However, in countries where assessments have been in operation longest (notably, the UK), the benefits listed in Table I are real and substantial. For a while, benefits probably do exceed costs. It is significant that more countries have introduced performance-based systems to allocate at least some of their resources. Yet, even if Figure 1 reflects the relative benefits and costs better than Figure 2, at a certain time, the benefits are likely to fall below the costs again. For a while, benefits in the UK may well have risen above costs, but the benefit curve has probably peaked, and diminishing returns have perhaps set in.[88] One can only speculate whether the benefit curve has yet fallen below the cost curve. However, support for this comes from suggestions that the 2001 RAE may be the last of its kind – or that if it is kept, it will be re-designed so as to reduce its expense.[89]

Given the substantial (and perhaps increasing) costs of a fully performance-based system, it is worth pointing to the advantages of a 'hybrid' system, based partly on performance (incentive-creating) and partly on educational size (cost-minimizing). The Netherlands, Finland, and Denmark have such systems. In addition, there is a danger with assessments like the British RAE that focus upon a one-dimensional concept of quality, and which link the results directly to funding. In The Netherlands, by contrast, peer review is used to assess performance in four dimensions – scientific quality, scientific productivity, scientific relevance, and long-term viability – and rankings are not directly linked to funding. The greater breadth of this quality measurement encourages greater diversity, which is undoubtedly an advantage in helping to ensure the 'health' of a nation's academic research.[90] The absence of any direct connection between funding and output generates many of the advantages identified in

[87] Some suggest that this is the situation in regard to the Teaching Quality Assessments, to which UK universities are also subject.

[88] See note 84.

[89] See Thomson, *op. cit.* note 80; A. Goddard, 'Hands-Off Approach for Top Research', *The Times Higher Education Supplement*, 1431, 14 April 2000, 68; and R. Floud, 'Universities Are Sinking Under Inspection Load', *The Times Higher Education Supplement*, 1479, 23 March 2001, 16.

[90] Cf. Royal Academy of Engineering, *Measuring Excellence in Engineering Research* (London: Royal Academy of Engineering, 2000).

Table I (e.g., protecting autonomy, encouraging the integration of teaching and research). For countries contemplating a performance-based system, such a hybrid approach may offer better prospects for the future, than the more expensive methods used in the UK.

ACKNOWLEDGEMENTS

This project was supported by the Higher Education Funding Council of England. This paper derives from a report by A. Geuna, D. Hidayat, and B. Martin, *Resource Allocation and Research Performance: The Assessment of Research* (Brighton: SPRU, 1999). The authors are extremely grateful to Dudi Hidayat who provided research assistance, and who co-authored the report. They would also like to thank Ammon Salter, as well as the Editor of *Minerva*, and three anonymous referees for their comments.

ABOUT THE AUTHORS

Aldo Geuna is a Senior Lecturer at SPRU, University of Sussex. He has been a Research Fellow at BETA, Université Louis Pasteur (Strasbourg 1), and at MERIT, Maastricht University. He obtained his PhD from Maastricht University. His research interests include the economics of knowledge production and distribution, science and technology policy, and the economics of innovation and technological change. He has written *The Economics of Knowledge Production* (Edward Elgar, 1999) and has co-edited *Science and Innovation: Rethinking the Rationales for Funding and Governance* (Edward Elgar, 2003).

Ben Martin is Professor in Science and Technology Policy Studies, and Director of SPRU (Science and Technology Policy Research) at the University of Sussex. With John Irvine, he helped establish techniques for the evaluation of scientific laboratories, research programmes, and national science. He also pioneered the concept of 'foresight' as a policy instrument, and served as a member of the Steering Group for the UK Foresight Programme from 1993 to 2000. His recent work includes an analysis of the benefits from government funding of basic research.

SPRU – Science and Technology Policy Research
The Freeman Centre
University of Sussex
Falmer, Brighton, Sussex BN1 9QE
UK
E-mail: A.Geuna@sussex.ac.uk