Performance-based Funding for Public Research in Tertiary Education Institutions

WORKSHOP PROCEEDINGS



OECD

Performance-based Funding for Public Research in Tertiary Education Institutions

WORKSHOP PROCEEDINGS



This work is published on the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Organisation or of the governments of its member countries.

Please cite this publication as:
OECD (2010), *Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings*, OECD Publishing.
http://dx.doi.org/10.1787/9789264094611-en

ISBN 978-92-64-09460-4 (print)
ISBN 978-92-64-09461-1 (PDF)

Corrigenda to OECD publications may be found on line at: *www.oecd.org/publishing/corrigenda*.

© OECD 2010

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to *rights@oecd.org*. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at *info@copyright.com* or the Centre français d'exploitation du droit de copie (CFC) at *contact@cfcopies.com*.

Foreword

This publication presents a collection of papers from the OECD-Norway workshop on performance-based funding of public research in tertiary education institutions, held in Paris on 21 June 2010. The workshop was organised by the OECD Directorate for Science, Technology and Industry (DSTI) in conjunction with the Norwegian delegation to the OECD. It brought together delegates from the OECD Working Party on Research Institutions and Human Resources (RIHR) with experts and policy makers from OECD countries and non-member economies.

The workshop formed part of the RIHR group's ongoing analysis of the governance and funding of research-performing institutions. Governments are seeking to channel research funds to universities and other institutions in ways that support quality research in strategically important areas and bolster the effective diffusion of knowledge. The issues of steering and funding take on greater relevance in light of the financial crisis and economic downturn, which have put severe fiscal pressure on many countries. The workshop was an opportunity to take stock of current thinking and practice about performance-based funding of public research in tertiary education institutions, as one way to help governments meet their research goals. These funding models essentially involve *ex post* evaluation of research outputs and outcomes from universities and other tertiary institutions, and are generally based on peer review, bibliometric or other quantitative indicators. The results are used to inform government decisions about how much and which institutions to fund.

The project was managed by Sarah Box of the OECD DSTI's Science and Technology Policy Division. It benefited from the leadership of the steering group, comprising representatives from Norway (lead country), Denmark, France, Germany, Poland, Sweden and the United Kingdom. The OECD Secretariat would also like to acknowledge the financial assistance provided by Belgium, Denmark, France, Norway and Sweden, and to thank all participating countries for their comments and contributions to the work.

Table of contents

Executive summary 9

Résumé *15*

Chapter 1. Overview of models of performance-based research funding systems 23

Introduction 24
Coverage 25
Rationales 27
The features of PRFSs 31
Funding and the PRFS 38
Convergence and challenges 42
Conclusion 46
References 49

Chapter 2. Performance indicators used in performance-based research funding systems 53

Introduction 54
The concept of performance and the rich world of indicators 55
The construction of national PRFSs 66
Consistency in the measurement of performance? 76
Notes 80
References 81

Chapter 3. Performance-based funding for public research in tertiary education institutions: Country experiences 85

Introduction 86
Institutional and project funding – the backdrop 86
Systems of performance-based funding 88
Key findings and policy issues raised by the questionnaire responses 121
Acronyms 123
Notes 125
References 126

Chapter 4. Impacts of performance-based research funding systems: A review of the concerns and the evidence..... 127

Introduction 128
Intended consequences 131
Impact.....133
Additional comments 154
Conclusion.....158
Acronyms 160
Notes 161
References 161

Chapter 5. Highlights and reflections: Rapporteur's report..... 167

The changing role of universities 168
Performance-based research funding systems: a concept with many variations 169
Some policy issues 172
A possible research agenda to develop more robust PRFS 172
Notes 174
References 174

Annex A: OECD-Norway workshop on performance-based funding for public research in tertiary education institutions: Workshop summary .. 175

Main discussion themes 183

Tables

Table 1.1. National performance-based research funding systems for universities 26
Table 2.1. Input indicators 57
Table 2.2. Process indicators.....58
Table 2.3. Structure indicators 59
Table 2.4. Output indicators.....60
Table 2.5. Effect indicators 60
Table 2.6. National performance-based funding systems: Historical background .. 66
Table 2.7. PRFSs using first-order indicators monitoring input and output.....69
Table 2.8. PRFS using first-order indicator monitoring effects 70
Table 2.9. PRFSs using third-order indicators 73
Table 2.10. Data sources and field differentiation 76
Table 2.11. Characteristics of good indicators.....77
Table 3.1. Key features of performance-based funding schemes in operation..... 95

Figures

Figure 1.1. PRFSs classified based on use of bibliometrics 36
Figure 2.1. A systemic framework for organisational performance 55
Figure 3.1. Government funded R&D in higher education, by type of funding, 2008 87
Figure 4.1. Australian university responses to government funding initiatives .. 139
Figure 4.2. Data-based predicted evolution of Spanish scientific production from 1974 to 1990 140
Figure 4.3. The relative citation impact of the UK research base, 1981-2007 142
Figure 4.4. Change in publication outlets by major field of research, 2003 to 2008 152

Boxes

Box 3.1. Funding research in Slovenia 92
Box 3.2. Rewarding research activities: Polish feedback 113

Executive summary

Policy interest in performance-based funding for research is growing

Tertiary education institutions (TEIs) are a key pillar of countries' research and innovation systems. As part of the public research system, they play many roles: education, training, skills development, problem solving, creation and diffusion of knowledge, development of new instrumentation, and storage and transmission of knowledge. Expenditure on R&D performed by the higher education sector forms a relatively small component of total R&D spending in OECD countries: an average of 0.4% of gross domestic product (GDP) in 2008, equivalent to around 17% of total gross domestic expenditure on R&D. Nevertheless, because TEIs often undertake longer-term, higher-risk research activities at some distance from commercialisation, they make an essential contribution to the knowledge base and complement the research activities undertaken by the private sector.

Given the importance of TEIs for countries' research and innovation performance, governments must think seriously about how best to arrange their governance and funding. Performance-based research funding is a relatively recent policy tool (which has been in use somewhat longer in the United Kingdom). It seeks to channel research funding according to *ex post* assessment of institutions' research outputs and outcomes. The OECD's Working Party on Research Institutions and Human Resources (RIHR) expressed a strong interest in mutual learning based on experience with performance-based funding in universities and other TEIs. Of particular interest was understanding more about the positive and negative impacts of such funding systems and how they dealt with institutional diversity in terms of disciplines, institutional size and budgets.

To begin to answer some of these questions, the OECD Secretariat established a steering group, which devised a project plan with four blocks of work that aimed to shed light on these issues. Experts were commissioned to undertake investigations of models, indicators and impacts of performance-based funding schemes. In addition, the RIHR conducted a country survey of models and schemes used. The outputs from this work formed the basis of the OECD-Norway workshop that took place in Paris on 21 June 2010. This volume brings together the papers prepared for the workshop as well as a rapporteur's report and workshop summary.

Current funding models have numerous goals, including research excellence

Diana Hicks' work on the models of performance-based research funding systems (PRFSs) (Chapter 1) draws on literature on both higher education and innovation to explore the rationales, features, funding and future of these increasingly popular evaluation systems. Thirteen countries were identified as using systems of *ex post* research output evaluation, with evaluation results influencing the distribution of government funding. Their systems were motivated by several themes, including resource concentration, encouraging international publication, and the general pursuit of excellence. In general, the methods used in assessment tend to correlate with the choice of focus; peer review and peer judgment based on indicators are used for individual and departmental evaluations, while quantitative formulas are used for university-level evaluations. The direct and indirect costs of assessments can be large, but these are rarely discussed in the literature. In practice, the tension between complexity and practicality means that while research groups are theoretically the ideal unit of evaluation, departments or universities are usually the focus of PRFSs. Ongoing challenges are to adequately assess the output of social science and humanities fields and to represent and recognise the broader impact of research.

The percentage of funding that depends on the research evaluation is a key feature of PRFSs, but the relevant definitions differ and information is limited. While the amounts of money involved may be small, they can have strong incentive effects, particularly if results affect institutional prestige or entrain other parts of the research funding system. In addition, the influence of a PRFS will depend on how institutions allocate funds internally; this is affected by their level of autonomy as well as internal governance arrangements. Looking ahead, the future of PRFSs depends on how successful they prove to be compared to other mechanisms and how well they meet governments' goals. Since contestability seems to be at the heart of the benefits of a PRFS, it may be possible to encourage research excellence through other means, such as rankings or centre-of-excellence block grants. However, seeking a broader set of goals may lead to conflicts, and increasing the complexity of PRFSs to deal with a wider definition of performance will increase costs. A suite of programmes to serve different goals would increase diversity in the funding system and facilitate differentiation among institutions, and should enhance the ability of the system as a whole to serve a complex array of public values.

Performance indicators are varied and constantly evolving

Hanne Foss Hansen's discussion of performance indicators used in PRFSs (Chapter 2) notes that while performance-based management has become widespread in the public and not-for-profit sectors, the concept of performance remains ambiguous. There is an ongoing discussion in the academic community about the definition of quality in research outputs and the gap between outputs and their effects. Against this backdrop, Hansen identifies three groups of indicators and analyses their strengths and weaknesses. First-order indicators, related to inputs, processes, structures, outputs and effects, include variables such as external funding, numbers of students, publications and knowledge transfer. All have some limitations; for example, citation counts may relate as much to communication structures and professional networks as to quality, and deriving and interpreting citation data raises technical challenges. As a reaction to methodological issues related to citation analyses, second-order indicators that take the form of summary indexes have been developed. However, problems with these mean they should be used with caution. Third-order indicators, which revolve around peer review, are an important mechanism for quality control and resource distribution at the micro level in the research community. Over time, different forms have developed, sometimes supported by first and second-order indicators. Importantly, all these measures are proxies of performance, and ongoing debates regarding their strength have led to continuing indicator development. In addition, the integration of innovation policy with research policy has given rise to new indicators related to knowledge transfer and commercialisation.

PRFSs have depended on either third-order indicator models mainly based on peer review or first-order indicator models mainly based on monitoring input and output of research institutions. At present a third model seems to be gaining ground: a first-order indicator model that monitors effects and is anchored in the idea of counting citations. There appears to be correlation between citation analysis and peer review in terms of their results, although this varies among disciplines. There is a need to look more closely into this correlation and also into how peer reviews are carried out in PRFSs. More generally, more work is needed on the "performance paradox", that is, weak correlation between indicators and performance itself. There is a knowledge gap concerning whether and how PRFS indicators deteriorate over time and how they incentivise academics. Contexts and actors' strategies shape and are shaped by PRFSs and more work is needed to understand these dynamics.

Budgetary impacts also differ across countries

Chapter 3 by Sarah Box presents the results of an OECD RIHR country survey on funding models. It outlines the key features and effects of performance-based funding systems currently in use. Most current schemes have been introduced since 2000, and funding generally operates on an annual basis, although some assessment processes are conducted less frequently. A third of schemes covers only public TEIs, but all cover all types of research and fields of activity. The overall range of indicators used by countries is similar, but various combinations and weightings are employed. The dimensions that differ most, in addition to the indicators, are the budget impact of the schemes and the role of institutions in their development and administration. Over time, indicators have been fine-tuned, but criticisms remain, particularly regarding their relevance, field-specificity and reward structure.

There have been few formal evaluations of PRFSs, but the evidence suggests positive effects on research outputs and research management. However, it is difficult to attribute these effects directly to the schemes, as the funding environment is multifaceted. Negative and unintended effects were also highlighted, such as a narrowing of research focus. Some interesting policy issues emerged from the questionnaire results, including whether it is effective to have multiple goals for PRFSs, which indicators have the greatest impact on research incentives, what level of budget impact is required to provide incentives to TEIs, and how government co-ordination mechanisms can be arranged to avoid friction between different funding tools and channels.

Evidence-based analysis of impacts is scarce

Linda Butler's work on the impacts of performance-based research funding systems (Chapter 4) attempts to highlight the unintended consequences of these systems. With a relative scarcity of evidence-based analysis in the literature, it seeks to distinguish between reality and perception and between evidence-based and anecdotal claims. The impact of a PRFS is multifaceted and a number of actors influence final responses – institutions, managers and researchers themselves. Intervening factors can also mitigate expected outcomes. Butler explores the available evidence on funding, human resource issues, productivity, quality, teaching, discipline mix, the focus of research, collaboration, institutional management practices, publishing practices and author behaviour. The review raises questions about the extent to which anecdotal evidence or survey responses can provide an accurate picture of

behavioural responses and Butler calls for a co-ordinated study of the impact of PRFSs on the academic endeavour.

Unsurprisingly, the strategies adopted by universities in response to PRFSs seek to maximise the returns (funding) they receive. Some practices have been criticised, although others, such as the "transfer market" in top researchers prior to evaluation rounds, might also be seen as rewarding high achievers. The response to assessment differs more among institutions; a poor assessment may lead one university to close a department, while another may decide to invest more resources to improve outcomes in that department. Importantly, PRFSs do not operate in isolation and other funding mechanisms as well as unrelated government policies can temper the consequences of evaluations. At the same time, the introduction of a PRFS can have an impact beyond the higher education or research sector, as research involves a myriad of inter-linked stakeholders. Better understanding the impacts of PRFSs would allow policy makers to assess the efficacy of their use and develop strategies to overcome adverse consequences.

International comparative studies of impact would aid policy making

The workshop rapporteur, Marc Luwel, outlined the main topic discussed at the workshop, identified some key policy issues and suggested possible next steps for the work on PRFSs (Chapter 5). He noted the changing role of universities and the increasing use of performance-based management and market-type incentives to steer their behaviour. While several countries have introduced PRFSs, there is no "ideal" methodology and there is intense debate about indicators and the intended and unintended consequences of systems. The objectives of transparency and accountability have been achieved, and there has been an improvement in the information management systems of universities. However, claims and counterclaims continue regarding game playing, citations, risk averseness and research collaboration. Over the next few years, universities will be confronted with challenges on the content of their "social contract" and will also face pressures to do more with less. International comparative studies of the impacts of PRFSs that draw on in-depth knowledge of national systems could be of great benefit to national authorities and universities in their efforts to increase the effectiveness and efficiency of institutional funding.

Finally, Annex A provides an overview of the presentations and discussions that took place at the workshop, highlighting the main themes of debate and areas for which participants felt further analysis and exploration would be valuable. Participants agreed that international comparative studies would be valuable, particularly those that could identify general developments affecting all countries as well as specific country trends.

Résumé

Les pouvoirs publics s'intéressent de plus en plus au financement de la recherche fondé sur les résultats

Les établissements d'enseignement supérieur constituent l'un des piliers du système de recherche et d'innovation d'un pays. Dans le cadre du système de recherche public, ils jouent de nombreux rôles : enseignement, formation, développement des compétences, résolution de problèmes, création et diffusion de savoir, mise au point de nouveaux instruments, stockage et transmission des connaissances. Les dépenses consacrées à la R-D réalisée dans l'enseignement supérieur représentent une part relativement limitée des dépenses totales de R-D des pays de l'OCDE, soit en moyenne 0.4 % du produit intérieur brut (PIB) en 2008, l'équivalent d'environ 17 % de la dépense intérieure brute totale de R-D. Néanmoins, étant donné que les établissements d'enseignement supérieur sont souvent engagés dans des activités de recherche à long terme, qui comportent davantage de risques et sont relativement éloignées du stade de commercialisation, ils apportent une contribution essentielle à la base de connaissances et complètent les activités de recherche du secteur privé.

Étant donné l'importance que revêtent les établissements d'enseignement supérieur pour les performances des pays en matière de recherche et d'innovation, les pouvoirs publics doivent mener une réflexion sérieuse sur les meilleurs moyens d'en assurer la gouvernance et le financement. Le financement de la recherche fondé sur les résultats est un instrument relativement récent (en usage depuis un peu plus longtemps au Royaume-Uni), qui vise à orienter le financement de la recherche en fonction d'une évaluation *ex post* des résultats et produits de la recherche réalisée par les établissements. Le Groupe de travail de l'OCDE sur les institutions et les ressources humaines de la recherche a manifesté un grand intérêt pour l'apprentissage mutuel à partir de l'expérience de ce type de financement dans les universités et autres établissements d'enseignement supérieur. Il s'agissait notamment de mieux comprendre les incidences positives et négatives des systèmes de financement fondé sur les résultats et les modalités de leur application à la diversité institutionnelle en termes de disciplines, de tailles d'établissements et de budgets.

Pour commencer à répondre à certaines de ces questions, le Secrétariat de l'OCDE a créé un groupe de pilotage qui a élaboré un plan de projet articulé sur quatre modules de travail. Des experts ont été chargés de mener des études sur les modèles, les indicateurs et les impacts des systèmes de financement fondé sur les résultats. Le Groupe de travail sur les institutions et les ressources humaines de la recherche a en outre mené une enquête par pays sur les modèles et systèmes en place. Les résultats de cette enquête ont servi de point de départ de l'atelier OCDE-Norvège qui s'est tenu à Paris le 21 juin 2010. La présente publication rassemble les documents établis pour cet atelier ainsi que le compte rendu du rapporteur et un résumé de l'atelier.

Les modèles actuels de financement visent de nombreux objectifs, notamment l'excellence en matière de recherche

L'étude de Diana Hicks sur les modèles de systèmes de financement de la recherche fondé sur les résultats (chapitre 1er) s'appuie sur une documentation concernant l'enseignement supérieur et l'innovation pour examiner la logique, les caractéristiques, le financement et les perspectives d'avenir de ces systèmes d'évaluation de plus en plus populaires. L'auteur a ainsi recensé 13 pays utilisant des systèmes d'évaluation *ex post* des produits de la recherche et dans lesquels les résultats de cette évaluation influençaient la distribution du financement public. Ces systèmes étaient axés sur plusieurs objectifs, notamment la concentration des ressources, l'encouragement à la publication internationale et la recherche générale de l'excellence. Globalement, les méthodes utilisées dans l'évaluation ont un lien avec l'objet de l'évaluation. Ainsi, les examens par les pairs fondés sur des indicateurs sont utilisés pour les évaluations individuelles et par département, tandis que les méthodes quantitatives sont privilégiées pour l'évaluation au niveau de l'établissement. Les coûts directs et indirects des évaluations peuvent être considérables, mais on ne trouve guère d'études qui leur soient consacrées. En pratique, la tension entre complexité et possibilité d'application pratique implique que même si le groupe de recherche constitue en théorie l'unité d'évaluation idéale, les systèmes de financement de la recherche fondé sur les résultats sont en général appliqués au niveau des départements ou des universités. La difficulté qui se pose constamment consiste à mesurer correctement les résultats de la recherche dans les domaines des sciences sociales et humaines, et de représenter et prendre en compte leur impact plus large.

La proportion du financement qui dépend de l'évaluation de la recherche est un élément clé des systèmes de financement fondé sur les résultats, mais les définitions varient et l'information à cet égard est limitée. Les sommes peuvent être faibles, mais avoir de puissants effets incitatifs, en particulier si

les résultats influent sur le prestige de l'établissement ou ont un effet d'entraînement sur d'autres parties du système de financement de la recherche. En outre, l'influence d'un système de financement fondé sur les résultats dépendra des modalités d'attribution des fonds à l'intérieur des établissements, lesquelles sont fonction du degré d'autonomie ainsi que du mode de gouvernance interne des établissements. L'avenir de ce type de systèmes sera déterminé par leur efficacité par rapport à d'autres mécanismes et par leur capacité à répondre aux objectifs des pouvoirs publics. Étant donné que la contestatibilité semble au cœur des avantages d'un tel système, il est peut-être possible d'encourager l'excellence dans la recherche par d'autres moyens, par exemple le classement ou des subventions globales à des centres d'excellence. Cependant, si l'on vise une série d'objectifs plus larges, on s'expose à des conflits de priorité, et si l'on complexifie les systèmes de financement de la recherche fondé sur les résultats pour élargir la définition des résultats, les coûts augmenteront. Une série de programmes axés sur des objectifs différents favoriserait la diversité du système de financement et la différenciation entre établissements, et devrait renforcer la capacité du système dans son ensemble à servir un éventail complexe de valeurs publiques.

Les indicateurs de résultats sont variés et ne cessent d'évoluer

Dans son examen des indicateurs de résultats utilisés dans les systèmes de financement de la recherche fondé sur les résultats (chapitre 2), Hanne Foss Hansen observe que si la gestion axée sur les résultats s'est largement répandue dans les secteurs publics et sans but lucratif, le concept de résultats demeure ambigu. La définition de la qualité des résultats de recherche et l'écart entre les résultats et leurs effets font débat dans les milieux universitaires. Dans ce contexte, Hansen met en évidence trois groupes d'indicateurs et en analyse les avantages et inconvénients. Les indicateurs du premier ordre, qui concernent les intrants, les processus, les structures, les résultats et les effets, comprennent des variables telles que le financement externe, l'effectif d'étudiants, les publications et le transfert de connaissances. Tous ces indicateurs ont leurs limites. Par exemple, le nombre de citations peut autant être lié aux structures de communication et aux réseaux professionnels qu'à la qualité, et l'interprétation des données qui s'y rapportent pose des difficultés techniques. Pour remédier aux difficultés méthodologiques liées à l'analyse des citations, on a élaboré des indicateurs du deuxième ordre qui se présentent sous forme d'index sommaires. Cependant, les problèmes qu'ils posent obligent à la prudence. Les indicateurs du troisième ordre, qui concernent l'examen par les pairs, sont un important mécanisme de contrôle de la qualité et de distribution des ressources au niveau local dans le monde

de la recherche. Avec le temps, différentes formes ont vu le jour et ces indicateurs sont parfois complétés par des indicateurs du premier ou du deuxième ordre. Chose importante, toutes ces mesures sont des variables indicatives des résultats, et le débat permanent sur leur solidité a incité à poursuivre les travaux sur les indicateurs. En outre, l'intégration de la politique de l'innovation à celle de la recherche a donné naissance à de nouveaux indicateurs concernant le transfert du savoir et la commercialisation.

Les systèmes de financement de la recherche fondé sur les résultats dépendent soit des modèles d'indicateurs du troisième ordre, qui reposent essentiellement sur l'examen par les pairs, soit de modèles d'indicateurs du premier ordre, qui mesurent les intrants et la production des établissements de recherche. Actuellement, un troisième modèle semble gagner en importance : c'est un modèle d'indicateurs du premier ordre qui suit l'effet des résultats de recherche et intègre le comptage des citations. Il semble exister une corrélation entre l'analyse des citations et l'examen par les pairs en termes de résultats, avec toutefois des différences selon les disciplines. Il importe d'examiner plus attentivement cette corrélation et de déterminer comment les examens par les pairs sont effectués dans les systèmes de financement fondé sur les résultats. Plus généralement, il faut pousser la réflexion sur le « paradoxe des résultats », ou l'existence d'une faible corrélation entre les indicateurs et les résultats proprement dits. On ne sait pas bien encore si les indicateurs des systèmes de financement fondé sur les résultats se dégradent avec le temps, et le cas échéant comment, ni comment ils motivent les universitaires. Il existe une influence réciproque entre, d'une part, le contexte et les stratégies des acteurs concernés et, d'autre part, le système de financement. La réflexion doit se poursuivre pour comprendre cette dynamique.

Les impacts budgétaires varient eux aussi selon pays

Le chapitre 3 par Sarah Box présente les résultats d'une enquête par pays sur les modèles de financement, réalisée par le Groupe de travail de l'OCDE sur les institutions et les ressources humaines de la recherche. Il décrit les caractéristiques et les effets les plus importants des systèmes de financement fondé sur les résultats actuellement en place. La plupart de ces systèmes, introduits depuis 2000, fonctionnent sur une base annuelle, bien que certains processus d'évaluation soient menés à intervalles plus espacés. Un tiers des systèmes ne sont appliqués qu'aux établissements d'enseignement supérieur publics, mais tous couvrent tous les types de recherche et champs d'activité. L'éventail global d'indicateurs utilisés est sensiblement le même dans tous les pays, mais les comparaisons et pondérations peuvent être très différentes. Indépendamment des indicateurs, les dimensions qui

varient le plus sont l'impact budgétaire des systèmes et le rôle des institutions dans leur développement et leur administration. Les indicateurs ont été affinés avec le temps, mais des critiques subsistent, en particulier en ce qui concerne leur pertinence, leur spécificité par rapport au champ de recherche et les structures de rémunération.

Les systèmes de financement de la recherche fondé sur les résultats n'ont guère fait l'objet d'évaluations en bonne et due forme, mais certaines indications portent à croire qu'ils peuvent avoir des effets positifs sur les résultats et la gestion de la recherche. Cependant, il est difficile d'attribuer ces effets directement aux systèmes proprement dits, car de nombreux facteurs interviennent dans le financement. Des effets négatifs ou indésirables ont également été mis en évidence, notamment le rétrécissement du champ de la recherche. Les réponses au questionnaire ont fait apparaître certaines questions de fond intéressantes, notamment celles de savoir s'il est efficace que les systèmes de financement fondé sur les résultats visent de multiples objectifs, quels indicateurs ont l'impact incitatif le plus puissant sur la recherche, quel niveau d'impact budgétaire est nécessaire pour motiver les établissements d'enseignement supérieur, et enfin comment les mécanismes de coordination publics peuvent être conçus pour éviter la tension entre différents instruments et voies de financement.

L'analyse des impacts à partir de données probantes est peu répandue

Les travaux de Linda Butler sur l'impact des systèmes de financement de la recherche fondé sur les résultats (chapitre 4) visent à mettre en lumière les conséquences indésirables de ces systèmes. Compte tenu de la relative rareté d'analyses étayées par des données probantes, l'auteur tente de cerner la distinction entre la réalité et la perception, et entre les arguments reposant sur des faits et les observations à caractère ponctuel. L'impact des systèmes de financement de la recherche fondé sur les résultats comporte divers aspects et peut faire intervenir plusieurs acteurs – établissements, gestionnaires et chercheurs eux-mêmes. Les facteurs qui entrent en jeu peuvent également atténuer les résultats attendus. Butler examine les données disponibles sur le financement, les questions de ressources humaines, la productivité, la qualité, l'enseignement, l'éventail de disciplines, les domaines de recherche privilégiés, la collaboration, les pratiques en matière de gestion institutionnelle et de publications, et le comportement des auteurs. Elle soulève la question de savoir dans quelle mesure des données à caractère ponctuel ou des réponses à des questionnaires peuvent permettre de brosser un tableau exact de l'évolution des comportements. Butler préconise une étude coordonnée de l'impact que les systèmes de financement fondé sur les résultats peuvent avoir sur le dynamisme universitaire.

Comme on pouvait s'y attendre, les stratégies adoptées par les universitaires face aux systèmes de financement fondé sur les résultats visent à maximiser le rendement (le financement) qu'elles peuvent en tirer. Certaines pratiques à cet égard ont été critiquées, tandis que d'autres, telles que le « marché des transferts » des meilleurs chercheurs, qui se tient avant les séries d'évaluations, pourraient également être considérées comme un moyen de récompenser les chercheurs les plus performants. La réaction à l'évaluation est plus contrastée selon les établissements. En effet, une université faisant l'objet d'une évaluation médiocre pourrait être amenée à fermer un département, tandis qu'une autre pourrait décider au contraire d'investir davantage de ressources pour améliorer ses résultats dans ce même département. Chose importante, les systèmes de financement fondé sur les résultats ne fonctionnent pas en vase clos, et d'autres mécanismes de financement ainsi que des mesures indépendantes prises par les pouvoirs publics peuvent atténuer les conséquences des évaluations. Dans le même temps, l'introduction d'un système de ce type peut avoir un impact au-delà du secteur de l'enseignement supérieur ou de la recherche, car la recherche réunit de nombreux acteurs interdépendants. Une meilleure compréhension des impacts de ces systèmes permettrait aux décideurs d'évaluer l'efficacité de leur utilisation et d'élaborer des stratégies pour en éviter les conséquences défavorables.

Des comparaisons internationales de l'impact des systèmes seraient utiles aux décideurs

Le rapporteur de l'atelier, Marc Luwel, a décrit le thème principal de l'atelier, fait ressortir certaines questions de fond et proposé une voie à suivre pour les travaux sur les systèmes de financement fondé sur les résultats (chapitre 5). Il a ainsi noté l'évolution du rôle des universités ainsi que l'utilisation croissante de la gestion axée sur les résultats et des incitations de marché pour agir sur les comportements. Plusieurs pays ont mis en œuvre des systèmes de financement de ce type, mais il n'existe pas de méthodologie « idéale », et les indicateurs ainsi que les conséquences souhaitées ou indésirables de ces systèmes font l'objet d'un intense débat. Les objectifs de transparence et de responsabilité ont été atteints, et les systèmes de gestion de l'information des universités ont été améliorés. Cependant, arguments et contre-arguments continuent de s'opposer, notamment sur le fait que les établissements savent « jouer le jeu », sur la question des citations, l'aversion au risque et la collaboration dans la recherche. Dans les années à venir, les universités seront confrontées à des défis concernant le contenu de leur « contrat social » et subiront également des pressions pour faire davantage avec moins de ressources. Des études comparatives internationales de l'impact des systèmes de financement fondé

sur les résultats, s'appuyant sur une connaissance approfondie des systèmes nationaux, pourraient être très utiles aux autorités nationales et aux universités dans leurs efforts visant à accroître l'efficacité et l'efficience du financement institutionnel.

Enfin, l'annexe A donne un aperçu des exposés et des débats de l'atelier, en mettant en évidence les principaux thèmes du débat et les domaines sur lesquels les participants estimaient utile d'approfondir la réflexion et l'analyse. Les délégués sont convenus que des études comparatives internationales seraient très utiles, en particulier si elles permettent de mettre en évidence des évolutions générales affectant tous les pays ainsi que des tendances propres à certains d'entre eux.

Chapter 1

Overview of models of performance-based research funding systems

by

Diana Hicks
Georgia Institute of Technology

This chapter provides an overview of models of performance-based research funding systems. It discusses the rationales for such systems from a number of perspectives, including innovation studies and the higher education literature. It explores the variation in their design, the funding implications and the long-term future of these systems.

Introduction

The university research environment has been undergoing profound change. Performance-based research funding systems (PRFSs) are one of the many novelties introduced over the past few decades. This chapter explores the rationale and design of PRFSs and their relation to funding systems. It is based on a literature review conducted in early 2010 which sought to identify all known PRFSs and draw out as much detail on their operation as possible.

Understanding the literature on PRFSs requires acknowledging the dual identity of university research. On the one hand, university research is part of the larger enterprise of the university and is shaped by university governance and university-related policy making. On the other hand, university research is a substantial element of every national innovation system, and so is of concern to scholars of innovation and to governments seeking to enhance the innovativeness of their economies. These two perspectives on PRFSs differ somewhat. The higher education literature treats the research mission of universities in a somewhat sketchy fashion towards the end of documents that are mainly concerned with accreditation, completion rates, harmonisation, etc. The innovation literature tends to ignore the educational mission of universities and the changes under way in the allocation of the teaching component of university funding. This chapter draws from the higher education literature a framework for understanding the introduction of PRFS as part of new public management, a movement to reshape government. A detailed perspective on the methods used to assess research performance and their possible effects draws on the innovation literature.

The tension between autonomy and control recurs at several points in the discussion. Autonomy is a sensitive issue for scholars at both the institutional and individual levels. At least in the literature in English, critics often focus on the shifts of control over academic work that accompany the introduction of PRFS. Ultimately, the effects of PRFS on institutional autonomy seem ambiguous, and may become more so as extensive consultation with the academic community over PRFS design becomes routine. The effect of PRFSs on the perceived autonomy of individual scholars is a source of perennial dissatisfaction and of the accusation that PRFSs harm scholarship and the research enterprise.

This chapter is organised as follows. First the remit of the study is specified and a discussion of coverage lists the PRFSs. Then the rationales underpinning PRFSs are explored from several perspectives: the innovation and higher education literature and the stated rationales of governments. The

features of systems are explored as well as their funding implications. After examining the convergences between the systems, the future of PRFSs is considered.

Coverage

The scope of this study was specified by the OECD. These specifications were translated into the following list of criteria for systems to be included in the review:

- **Research** must be evaluated. Evaluations of the quality of degree programmes and teaching are excluded.
- Research evaluation must be ***ex post***. Evaluations of research proposals for project or programme funding are *ex ante* evaluations and are excluded.
- Research **output** must be evaluated. Systems that allocate funding based only on PhD student numbers and external research funding are excluded.
- Government distribution of research **funding** must depend on the results of the evaluation. *Ex post* evaluation of university research performance used only to provide feedback to the institutions is excluded.
- It must be a **national** system. University evaluations of their own research standing, even if used to inform internal funding distribution, are excluded.

The first PRFS was introduced in 1986 in the United Kingdom, the Research Assessment Exercise (RAE). Use of PRFSs has expanded since, and 13 countries were found to be using them in early 2010. These systems are listed in Table 1.1 which reports the countries, the name of the system or agency responsible, and the year the system was introduced. The systems are not static. Most have undergone major redesign. In fact new versions of seven systems are under discussion or in the first stages of rollout. In the PRFS literature, it is important to be aware of the version of the system that is discussed.

Table 1.1. National performance-based research funding systems for universities

Country	System	Year implemented/ major revision
United Kingdom	RAE moving to REF – research excellence framework	1986/current
Spain	CNEAI – National Commission for the Evaluation – *sexenio*	1989
Slovak Republic		1992/2002
Hong Kong, China	RAE	1993
Australia	Composite Index, Research Quality Framework (RQF), Excellence in Research for Australia (ERA)	1995/current
Poland	Ministry of Science and Higher Education – parametric evaluation	1991/1998-99
Italy	*Valutazione triennale della ricerca* (VTR)	Evaluation 2001-03/ funded 2009
New Zealand	Performance-based research funding (PBRF)	2003/current
Belgium (Flemish Community)	BOF key	2003/2008
Norway	Norwegian model (new model for result-based university research funding)	2006
Sweden	New model for allocation of resources	2009
Denmark	Implementation of the Norwegian model	Current
Finland	Funding formula for allocation of university resources	1998/2010

A number of countries do not have systems that meet the criteria for inclusion:

- France: In response to the Bologna process the French government is reworking its university evaluation mechanisms. Although this is driven by accreditation concerns, the research side appears to be included and the results influence funding distribution in some way.
- Germany: Although the *Länder* may be different, the federal government seems to be pursuing some goals of performance-based funding, such as international excellence, by awarding large centres of excellence to universities based on assessment of proposals (CHEPS, 2007, pp. 37-38).
- Japan: Funding for Japanese universities seems to be distributed based on number of faculty. Government attempts to introduce performance-based funding have met with opposition and have not yet succeeded.
- Korea: No known performance evaluation component in university funding. However, Seoul National University is becoming autonomous, and this sort of change is often accompanied by performance-based funding.

- Netherlands: The Netherlands implements an evaluation system based on peer review which is not used to inform distribution of funding (Geuna and Martin, 2003).
- Portugal: All research money appears to be project-based, although proposal evaluation may include *ex post* evaluation of the research team's performance (Strehl, 2007).
- South Africa: The Foundation for Research and Development conducts evaluations of individuals' research record. Peer-reviewed grants of quite lengthy duration are awarded based both on this *ex post* evaluation and on peer review of proposals.

Rationales

> *"The rationale of performance funding is that funds should flow to institutions where performance is manifest: 'performing' institutions should receive more income than lesser performing institutions, which would provide performers with a competitive edge and would stimulate less performing institutions to perform. Output should be rewarded, not input."* (Herbst, 2007, p. 90)

Straightforward as this seems, full understanding of the motivations behind PRFSs requires exploring some nuances. Broadly speaking, different parties evoke seemingly different types of rationales when explaining the introduction of national research evaluation systems. Those primarily concerned with understanding research and innovation tend to speak of globalisation, competitiveness, the knowledge economy, etc. Those who study higher education and who are mainly concerned with education, accreditation and the like tend to trace the introduction of these systems to the spread of the tenets of new public management. Discussions of the introduction and evolution of individual systems tend to take a historical turn, focusing on changes in government, ministerial actions and stated rationales of governments rather than on an interpretation of events in broader international contexts (Jiménez-Contreras *et al*., 2003; Sanz-Menendez, 1995; Tapper and Salter, 2003). Rationales of national governments tend to include the word "excellence". Of course, each approach addresses the same phenomenon from a different angle. For example, Kettl traces the widespread introduction of new public management reforms to a shared set of political, social, economic and institutional challenges faced by many governments. These include the shift from the industrial economy to the information economy and globalisation (Kettl, 2000).

An example of the global competitiveness argument can be found in the introduction to the recent European Commission report, *Assessing Europe's University-Based Research*:

> *"…research performance is widely regarded as being a major factor in economic performance. Because of their interlinked roles in education, research, and innovation, universities are considered key to the success of the Lisbon Strategy with its move towards a global and knowledge-based economy. Improving the capacity and quality of university-based research is thought to be vitally important for innovation, including social innovation…*
>
> *"The economic dimension of (university-based) research in terms of expected economic and societal benefit and increased expenditure goes a long way to explain the heightened concern for quality and excellence in research, for transparency, accountability, comparability and competition, and for performance indicators and assessment."* (European Commission, 2010, p. 9)

A very similar statement can be found on the UK government website introducing the new REF (Research Excellence Framework):

> *"Through the REF, the UK funding bodies aim to develop and sustain a dynamic and internationally competitive research sector that makes a major contribution to economic prosperity, national wellbeing and the expansion and dissemination of knowledge."* (HEFCE, 2010)

In contrast, higher education authors often view the changing relationship between universities and their government in the context of new public management reforms (for example, Herbst, 2007). Donald Kettl is a leading analyst of new public management, and in a nutshell his view is the following:

> *"Over the past quarter century, governments around the world have launched ambitious efforts to reform the way they manage. Citizens have demanded smaller, cheaper, more effective governments while asking for more programs and better services. To resolve this paradox, governments are experimenting with many ideas to improve performance and production and to reduce costs… Reviewing the standard strategies and tactics behind these reforms,* [Kettl's *The Global Public Management Revolution*] *identifies six common core ideas* (Brookings Institution Press, 2005):

– *greater productivity*
– *more public reliance on private markets*
– *a stronger service orientation*
– *devolution to sub-national government*
– *increased capacity to formulate and evaluate policy*
– *enhanced accountability.*"

Performance-based funding systems for public research in universities can be interpreted in light of these six core ideas (Kettl, 2000):

- Increasing productivity: Output-based evaluation increases research output without adding research resources to the system, thus increasing productivity (see Chapter 4).
- Replacing traditional command-and-control systems with market-like incentives: In many nations, universities were not autonomous and were often little more than extensions of their ministry. In Whitley's terms they were "administrative shells" (Whitley, 2008, p. 12), or as Herbst describes it "government agencies were practically distributing funds down to each individual faculty member… Until recently, we may claim in jest, a typical rector's office just had a yearly budget to pay for various banquets" (Herbst, 2007, p. 87). Universities had no discretionary budget and did not control hiring, tuition, student numbers, etc. The shift to performance-based funding is part of a broader movement to make universities more autonomous and introduce more strategic university management. This also involves competition for funding - the market-like side of the reform.
- Stronger service orientation: This dimension refers to increased attention to the needs of citizens. In the research world this would be analogous to giving more weight to serving the needs of the high-technology economy and less to the self-governed programmes of work of the community of scholars subject only to peer validation (Marginson, 1997, p. 69).
- Devolution: The idea here is that programmes are more responsive and effective when managed closer to the provision of services. In the university world this means making universities autonomous, strategically managed entities rather than having all key decisions taken by ministries.

- Formulating policy: This refers to a shift from formulating policy and delivering the service to formulating policy and contracting for the service. The government as purchaser of "education services" was explicitly articulated in the Australian context (Marginson, 1997, p. 71).
- Enhanced accountability: This means focusing on outputs and outcomes rather than processes and structures. Measuring research output and distributing funding on the basis of results is clearly meant to enhance accountability.

Not everyone views the application of new public management to universities in a neutral or positive light. Critical commentators see these developments in the context of neo-liberal policies, understood pejoratively. Also, although implementation of a PRFS is related to new public management, it is not an exemplar. Important new public management characteristics such as efficiency gains and increased use of contracting play little or no role in PRFSs.

The rationales provided by individual governments for their PRFSs are worth examining briefly because they reveal national differences in priorities due to the pre-existing state of university research in each country. In the United Kingdom, greater selectivity in funding allocation was initially an explicit goal of the RAE. Selectivity was a government response to limited resources and the increasing costs of research. The goal was to maintain research excellence but at fewer places (Tapper and Salter, 2003). Sweden also looks to concentrate resources, believing that international competition requires concentration and priorities to maintain high scientific quality (European Commission, 2010). The goal of the Spanish *sexenio* as stated in law is to foster university professors' research productivity and improve the diffusion of this research both nationally and internationally. Fostering international publication was an explicit goal of Spanish science and technology policy more generally, and overall the policies have been successful (Jiménez-Contreras *et al.*, 2003).

Excellence appears frequently as a goal. In Australia the stated goal of the new ERA system is to identify and promote excellence across the full spectrum of research activity, including discovery and applied research, in Australia's higher education institutions (ARC, 2009, p. 11). In New Zealand, the primary purpose of the PBRF is to ensure that excellent research in the tertiary education sector is encouraged and rewarded (New Zealand, Tertiary Education Commission, 2010). In Norway the goal is to increase research activities and allocate resources to centres performing excellent research (European Commission, 2010, p. 120). Sivertsen reports the goal of the Norwegian publication indicator to be "to measure and

stimulate the research activity at the level of institutions and to enhance the focus and priority they give to research as organisations" (Sivertsen, 2009, p. 6). Government statements of PRFS rationale thus reveal several independent themes: resource concentration, encouraging international (*i.e.* English-language) publication and the general pursuit of excellence. While the resource concentration theme bears some similarity to the search for increased efficiency of new public management reforms, research excellence is more reminiscent of the newer "public values" movement (Stoker, 2006).

The features of PRFSs

Performance-based research funding systems show variations in their design. Here their governance, unit of analysis, frequency, census period, cost, methods of measurement and use in funding allocation are examined.

There appear to be two basic models for running a PRFS; either it is run out of the Ministry of Education (or Ministry of Education and Research if there is no independent research ministry), sometimes based on a new law, or out of a new dedicated agency. The following is a list of economies and their PRFS administering agencies:

- Education ministries:
  - Hong Kong, China: University Grants Committee.
  - New Zealand: Tertiary Education Commission.
  - Slovak Republic: Act on Higher Education implemented by the Ministry of Education.
  - Sweden: Governmental Research Bill in 2008 established the new model for distribution of block grants system, handled by the Ministry of Education with some methodological support from the Swedish Research Council (European Commission, 2010).
  - United Kingdom: formerly Higher Education Funding Council for England (HEFCE), now Department of Business, Innovation and Skills.
- Mixed research/education ministries:
  - Australia: disputed between Department of Education, Employment and Workplace Relations and Department of Innovation, Science and Research.

  - Finland: Ministry of Education.
  - Norway: Ministry for Research and Education.
  - Poland: Science Council – advisory body to the Minister of Science and Higher Education (established by a new Act of October 8, 2004 on the Principles of Financing Science).

- Dedicated agencies:
  - Belgium (Flemish Community): Steunpunt O&O Statistieken (SOOS) agency established to produce the bibliometric analysis for BOF key, which is administered by the Flemish government.
  - Italy: currently Agency for the Evaluation of University System and Research (ANVUR), whose predecessor (CIVR) had an advisory relationship with the Council of Ministers and the Ministry of Universities and Research.
  - Spain: National Commission for Evaluation of Research Activity (CNEAI) established by law simply to implement the *sexenio*.

The choice of administrative home for the PRFS would seem to have nothing to do with the characteristics of the PRFS. Most likely the choice of education ministry or dedicated agency is a political decision or is based on structural characteristics of the government.

The possible units of research evaluation in PRFSs include: individuals, research groups, departments and universities. All have played a role in PRFSs. Note that the unit evaluated may or may not correspond to the unit allocated funding on the basis of the results. The Spanish *sexenio* and the New Zealand PBRF both grade individual researchers on their research record. The *sexenio* results are used to increase the salaries of individuals achieving a high grade. The PBRF results are aggregated into a rating for universities to be used in allocating block funding.

Research groups are in fact the unit of evaluation with the best theoretical rationale because research is conducted by such groups, not by individuals or departments. Departmental level PRFSs are routinely criticised because of this (see for example Herbst, 2007, p. 91). The recent assessment of university research evaluation in Europe by an expert working group of the European Commission makes a best practice recommendation that groups, or "knowledge clusters", be the preferred unit of analysis (European Commission, 2010, pp. 38-39). However, research group evaluation is probably impossible to implement on a national scale because there are so

many groups and their boundaries are fluid and indistinct. The Australian RQF proposed evaluation at the research group level. This elaborate exercise was the result of an extensive consultation and design exercise, but was abandoned by a new government before implementation in favour of the ERA, which emphasises simplicity. Because the ideal research group level evaluation is impractical, actual PRFSs will always be subject to criticism. However, departmental and university level PRFSs have proved their worth in practice and can be considered quite good and usable, even though imperfect.

The more practical and thus widely implemented unit of analysis is the department or the analogous field-in-university. The UK RAE evaluates university departments; the Australian ERA and Italian VTR evaluate fields in universities. In the RAE the results by department are published but are aggregated to the university level in order to award a single block grant to each university. The ERA is not yet used to allocate funding. In Poland faculties within universities are evaluated and money is awarded to faculties directly by the Ministry of Science and Higher Education. The Slovak Republic also appears to conduct its evaluation at the field-in-university level. Some systems produce evaluations at the university level. The Australian Composite Index, which is presently used to allocate the research quantum, is a formula that produces a result for each university. The same is true of the Norwegian and Danish PRFSs.

PRFSs also differ in the frequency with which evaluations are conducted. The Australian Composite Index and the Norwegian model require annual data submissions to calculate the following year's budget allocation. The Slovak Republic evaluates every three years (Geuna and Martin, 2003), Poland no less frequently than once every five years (in practice every four years), and Spain every six years (hence *sexenio*). The frequency of the Australian ERA has not yet been determined. Italy conducted an evaluation in 2006 using 2001-03 data and used the results to allocate funding in 2009 (European Commission, 2010, p. 116). Evaluations are not necessarily conducted on a regular schedule. The first two rounds of New Zealand's PBRF and Hong Kong, China's, RAE were conducted at three-year intervals. The next are scheduled at intervals of six and seven years, respectively. The interval between RAEs has increased from three to four, five and then seven years.

The census periods also differ among PRFSs. The Australian Composite Index and the Norwegian model base calculations on one year of data. The VTR was based on three years of data, Poland four years, Hong Kong, China, basically four years, the PBRF five years, the *sexenio* six years. ERA uses six years for bibliometric measures and three years for other indicators. RAE 2008 used publications from a seven-year period. The BOF key is

based on ten years for bibliometric measures and four years for non-bibliometric measures (see Poland, Ministry of Science and Higher Education, 2010; Hong Kong, China, UGC, 2006, p. 11; Hodder and Hodder, 2010; ARC, 2009, p. 13; Debackere and Glänzel, 2004, p. 268).

The costliness of the PRFS should be a concern for any government though cost is rarely discussed. PRFSs incur indirect costs for universities that compile submissions and direct costs for the evaluation of those submissions. The RAE and VTR incur heavy costs in faculty time because they are peer-review exercises. The RAE in 2001 involved 70 panels of 10 or more members convened to work on assessing 180 000 publications, which made the exercise expensive. Panels were expected to read papers, though given the impossibility of comprehensive reading, thoroughness varied (Harman, 2000, p. 115). One author noted that the exercise was conducted as if it was to appraise 50 000 individual researchers and their 180 000 pieces of work in order to make 160 funding decisions (Sastry and Bekhradnia, 2006), which seemed disproportionate. There were also indirect costs borne by departments whose efforts for preparing submissions increased with each round. A particularly clear cost statement was found for the Italian exercise which took place in 2006:

> *"The evaluation involved 20 disciplinary areas, 102 research structures, 18 500 research products and 6 661 peer reviewers (1 465 from abroad); it had a direct cost of EUR 3.55 million and a time length spanning over 18 months."* (Franceschet and Costantini, 2009, p. 1)

The VTR panels were expected to assess the quality of each of the 17 300 unique written works submitted, 2 900 of which were books. Two reviewers were assigned to assess each work independently. If every reviewer fully fulfilled their mandate, 5 800 reviewers read a book in addition to the four articles they were assigned. Is it any wonder that the intervals between the RAEs increased over time or that a second VTR has not been scheduled?

Indicator-based systems also incur substantial direct costs. The direct costs of indicator-based exercises include establishing and maintaining a national research documentation system, buying supplemental information from database providers, data cleaning and validation, and indicator calculation. As these data systems are intricate and large, the costs are considerable. Auditing of submissions is also necessary because fraudulent submissions could be used to increase funding allocations. Audits of Composite Index submissions were conducted by KPMG and found a high error rate (34% in the second audit in 1997); 97% of errors affected final scores and thus funding allocations (Harman, 2000, pp. 118-119). If costs are articulated, some consideration could be given to cost/benefit ratios.

The methods used in assessment correlate with the unit of analysis. Peer review is used for individual evaluations in Spain and New Zealand. Peer review is also used for departmental evaluations: the RAE, VTR and ERA. The RAE reviews are informed by narratives submitted by departments. In ERA, the reviewing committees are informed by indicators produced by an agency using university submissions. Although peer review seems to give control of the evaluation to the community of scholars, the criteria for reviewers vary. In Spain the criteria are set down in the regulations and are quite specific, reducing the degrees of freedom for reviewers and enhancing government control. In other countries less specific criteria provide scholars with more autonomy in their judgments (Jiménez-Contreras *et al.*, 2003). When the RAE began, disciplinary committees had complete freedom to determine criteria; later the government introduced a standardised statement of criteria to ensure fairness across fields (Tapper and Salter, 2003).

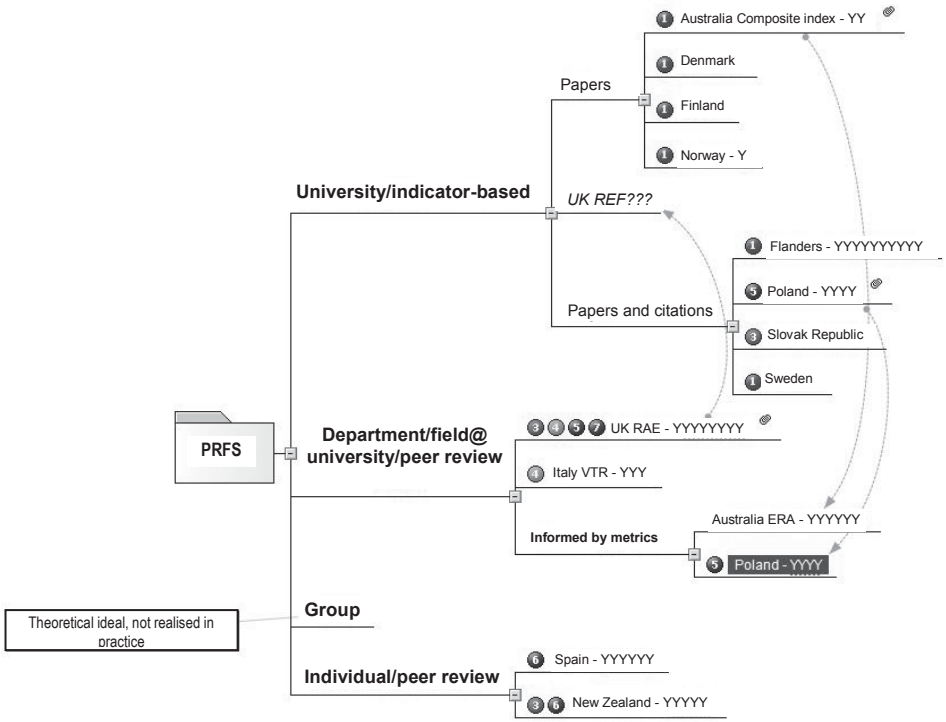
University-level evaluations rely on quantitative formulas. Such formulas use bibliometric output information but differ as to whether papers alone are counted or whether citation information is also included. The Australian Composite Index and the Norwegian and Danish systems employ paper counts only. However, Butler identified a weakness in simply counting papers, in that it encourages a move to lower quality journals to increase publication frequency (Butler, 2003). Therefore the Norwegian and Danish systems use an enhanced method in which 10% or more of journals are assigned to a higher quality category and given extra weight in the formula. Spain incorporates the Thomson-Reuters impact factor in their deliberations. Systems that include citation information, which must be purchased, include ERA, the Slovak Republic, Sweden and Belgium (Flemish Community). Spain's *sexenio* includes self-reported impact information such as citations or journal impact factor. The formulas include a range of other measures in addition to publication output including:

- employment of graduates,
- external research funding,
- faculty characteristics and qualifications,
- faculty size,
- graduate students graduated,
- implementation/application of research,
- international memberships,
- student load.

Figure 1.1 summarises this information in a broad typology of systems based on the unit evaluated/method used: individual/peer review, research group (theoretical, not realised), department or field in university/peer review, university/indicator based. A further subdivision based on type of bibliometrics used is included. The typology is simplified because non-bibliometric metrics are not considered, nor is the unit to which funding is allocated. However, if every dimension were included, every system would probably be in a different category since each would be unique.

Figure 1.1. PRFSs classified based on use of bibliometrics

Frequency, country and census period indicated



PRFSs are challenged by diversity. In the main this does not seem to concern the diversity of mission of universities (though Hong Kong, China, is concerned with this; French *et al*., 2001). PRFSs simply concern a segment of tertiary institutions, some of which obtain far more research money in the block grant than others. Since the goals of PRFSs revolve around encouraging excellence and sometimes international publishing or concentration of resources, they are not designed to help weak research institutions improve and may in fact serve to remove resources from teaching-oriented institutions. The national innovation systems literature is largely silent on this issue, presumably because the authors reside in research-intensive universities and are not particularly concerned with colleagues in teaching-oriented institutions who wish to compete for scarce research resources.

Treatment of field diversity is far more involved and all systems are sensitive to differences in the patterns of fields' output. Evaluating the social sciences and humanities presents particular challenges because traditional bibliometric evaluation using Web of Science data does not work well (Hicks, 2004; Hicks and Wang, 2009). Peer review systems convene field-based committees that generally have latitude for developing appropriate standards of judgment for their field (Poland, ERA, RAE, VTR). The ERA suite of quantitative indicators differs by field. The Swedish formula contains an intricate field-weighting system. Spain's *sexenio* permits submissions to include a broader range of publication types in social sciences and humanities, though there are indications that this does not work very well. The applications of social scientists are less likely to be approved than those of scientists, and success rates have not increased over time (Jiménez-Contreras *et al.*, 2003, p. 138).

Transparency is highly valued in the design, execution and distribution of the results of PRFSs. Most systems emphasise transparency of methods and data. Thus the systems are designed in highly consultative processes in which the government gathers input from universities using comment periods, and evaluation processes are designed by expert panels made up of representatives of university or field-based associations. Instructions to universities concerning their submissions are easily available over the Internet as are the formulas used to convert measures into final rankings, grades or weights. The final grades are publicly available as well, again often posted on government websites. In Norway and the RAE 2008, all university submissions are public. The public nature of the results means that they are often used by others, picked up by the media and used in other funding decisions such as internal university allocation of funds or as one factor considered in awarding research grants. Individual grades are used in tenure decisions in Spain and South Africa.

Funding and the PRFS

PRFSs are used in decision making concerning the distribution of block grants to universities. Older methods of deciding the distribution of block grants include using a historical basis, *i.e.* the previous year, or formulas based on faculty or student numbers. The introduction of research performance likely entails splitting the traditional block grant into teaching and research components. The percentage of funding that depends on the research evaluation is a key feature of the PRFS, but it is difficult to define and to obtain the necessary information in most cases. It can be defined in many ways, and there is no consistency among authors in this respect. Reported figures include shares of:

- total university resources (government plus other funding),
- government funding for universities (block grant plus research grants and contracts),
- block grant or "general university funds" (GUF),
- research resources (total or government).

The share allocated according to the performance formula might be of interest and might include research performance assessment as well as graduate student numbers and amount of outside funding raised, for example. A smaller share depending just on assessment of research output might be relevant. One might wish to know the average across all universities, the share only in the most research-intensive universities, or the highest and lowest shares in the system. In addition, static snapshots may be less interesting than trends over time. Finally, the amount that moves between universities in any two years might be most illuminating. Moreover, to fully understand the effects of PRFSs they must be compared with non-PRFS funding systems.

Ideally, each of these figures would be available for every university system for current and past allocations. The following limited information was found:

- In Australia, the Composite Index informs the distribution of DIISR research grants. In 2008 these grants comprised 10.5% of Australian government financial assistance to universities and 6% of total revenues from continuing operations at Australian universities (Australian Government Department of Education, Employment and Workplace Relations, 2009, Table 1).

- In Italy, 2% of block funding is allocated based on assessment of research output. This is calculated as follows: 7% of the block funding to universities (FFO, Fondo di Finanziamento Ordinario) is allocated according to a performance formula of which research comprises two-thirds of the total (and teaching one-third). Of this, 50% is in proportion to the grade received by the university from VTR in 2006 (European Commission, 2010, p. 115).
- In New Zealand, the research output evaluation comprises 60% of the PBRF, which is one of six Tertiary Education Organisation Component (TEOC) funding elements (Tertiary Education Commission, 2010). Thus the research output evaluation likely drives 10% of block grant funding in New Zealand.
- In Norway, the publication indicator only affects 2% of the total expenses of the higher education sector, and a publication point represents no more than EUR 5 000 (Sivertsen, 2009, p. 6).
- In the Slovak Republic in 2006, formula-based subsidies for research represented approximately 15% of the total budget (Strehl, 2007, p. 41).
- In the United Kingdom, the RAE informs the distribution of a block grant which comprised 36% of support for science and engineering R&D in UK universities in 2004-05 (Department of Trade and Industry [DTI], and Office of Science and Innovation, 2007). General support for university research is allocated largely, but not entirely, on the basis of RAE ratings. Other criteria such as supporting new subjects or making allowance for the higher cost of living in London also play a role. In 2006-07, HEFCE allocated 70% of its general support using RAE-linked criteria (HEFCE, 2006). Just prior to 2008 approximately 25% of all research support in UK universities appears to have been allocated to universities based on the RAE ratings of their departments. RAE-based allocations are quite stable. By 2001 the marginal impact of the RAE on university finances was small. Sastry and Bekhradnia (2006) calculated that as a result of the 2001 RAE, only one institution saw its total revenues affected by more than 3.7% and the median impact was less than 0.6%.

Some information on time trends is also available:

- In Belgium (Flemish Community), the block grant for universities is divided into three parts: general, basic research (BOF) and applied research (IOF), and 45% of general and BOF are allocated according to a formula called the BOF key. The weighting given to the evaluation

of research outputs in the calculation of the BOF key rose from 0.10 in 2003 to 0.30 in 2006 (Debackere and Glänzel, 2004, p. 268; Luwel, 2010).

- In Poland the proportion of research funding distributed in block grants is likely to decline as the government promises to increase the share dedicated to competitive awards (Pain, 2010).
- In the Slovak Republic, government funding for research in universities more than doubled between 2002 and 2005. At the same time, the share of block funding allocated according to evaluated research performance as well as the "quality of development projects" increased from 9.7% to 23.1% at the expense of historical allocation and allocation according to evaluated teaching performance (Strehl, 2007, p. 113).
- In the United Kingdom, the share of university resources from RAE-governed processes declined from 58% in 1984 to 35% in 1997 because the share of research funding from all other sources increased, including the share from competitive grants.

This information is difficult to summarise. The share of funding allocated using PRFSs ranges from almost trivial at 2% of total funding in Norway to quite substantial at perhaps 25% in the United Kingdom. Neither is the trend clear. In the United Kingdom the importance of RAE-based funding has decreased, while in Belgium (Flemish Community) and the Slovak Republic it seems to be increasing. However, analysts consistently emphasise the small amounts of money involved or the small amount that moves in any one year as a result of evaluation (Jiménez-Contreras *et al.*, 2003; Sanz-Menendez, 1995; Sivertsen, 2010; Sastry and Bekhradnia, 2006; Rodríguez-Navarro, 2009). Universities have high fixed costs and require stable funding streams, and no government would benefit from a university becoming bankrupt. Therefore, governments would not want a highly unstable funding system that would swiftly reallocate large amounts of funding and might bankrupt some institutions, though the Polish government is currently threatening to remove research funding from underperforming universities (Pain, 2010). The short-term financial consequences of PRFS are likely to be less than is often feared, although small, consistent movements can accumulate over time. Only the UK system has been operating long enough to assess the possible financial effects of PRFS over the long term. Since the UK government had an explicit goal of selectivity, or concentrating resources in fewer universities, the appearance of this effect should not have been unexpected.

PRFSs can create pressure to increase funding. Grade increases in the RAE between rounds were used to argue for more government money for research to reward improved performance (Tapper and Salter, 2003). Hong Kong, China, faces the same phenomenon (French *et al*., 2001). If the success rate in Spain's *sexenio* increases, the government is automatically obliged to increase faculty salaries. It seems likely that pressures to increase university funding will spread with the spread of PRFSs. Universities that can demonstrate increased excellence, particularly in highly publicised, internationally comparable, measurable ways, will have strong arguments when budgets are discussed.

It is possible that a PRFS will entrain other parts of the research funding system, with the result that the effect of the PRFS will go beyond the amount of money directly allocated. This will happen if grant review is not double-blind and the probability of a successful grant application is increased if the applicant is located in a higher-ranking department. In Spain a minimum number of *sexenios* are required for securing tenure or for becoming a member of the commission that grants tenure, thereby aligning the formation of an elite with the PRFS. Also, *sexenios* help with applications for competitive research funding by aligning the two major research funding mechanisms (Jiménez-Contreras *et al.*, 2003, p. 135).

Analysts who point to the small direct financial consequences attached to PRFSs do not therefore believe that they lack influence. On the contrary, they argue that the systems have strong effects on universities, less through the incentives funding provides than through public judgments about relative prestige. Comprehensive assessment of universities and their departments creates intense interest among universities. Experience has shown that universities are extremely responsive to hierarchical ranking. One effect of the RAE was to create what McNay termed assured, aspiring and anxious universities (HEFCE, 1997, p. 47). Attention devoted to RAE submissions did not decrease, even though, as mentioned above, Sastry and Bekhradnia calculated that the median impact on total university revenue of the 2001 exercise was 0.6%. Marginson noted in relation to the introduction of a university assessment in Australia in 1993:

> *"Nothing less than the positional status of every institution was at stake; the process of competitive ranking had a compelling effect, leading to the rapid spread of a reflective culture of continuous improvement."* (Marginson, 1997, p. 74)

Harman related that in Australia allocation of funding based on the Composite Index had become "an important vehicle for developing status hierarchies" as data are published in newspapers and widely used (Harman, 2000, p. 116).

The attention paid to research output assessments that are not linked directly to funding confirms this. In the United States, the annual ranking of university departments by *US News & World Report* is extremely influential. Similarly in Australia, ERA has been conducted but is not yet in use to determine funding allocation, yet it is the focus of intense interest. The Shanghai Jiao Tong University Academic Ranking of World Universities and the THE World University Rankings have been so influential that President Sarkozy has ordered France's science and higher education ministry to set "the objective of having two French establishments in the top 20, and 10 in the top 100" (Anonymous, 2010). That universities pay close attention to rankings and their attendant prestige is entirely rational since prospective students use rankings to decide on their destinations, especially at the graduate level, and money follows students. In addition, the more productive, grant-raising faculty seek to work at more highly ranked institutions.

Perhaps most tellingly, many UK universities may now be choosing high ranking over more money. RAE 2008 allowed selective inclusion of faculty members, and "research-intensive institutions indicated that they would seek the best ratings rather than the financial rewards that could be won by entering more staff" (Lipsett, 2007).

The emphasis in the literature on prestige rather than financial incentives as the main mechanism through which PRFSs work is consistent with conclusions of the new public management literature. Kettl (2000, p. 38), summarising conclusions of a report from the OECD's Public Management Committee, explains that "public managers around the world have indeed been strongly motivated by incentives, but the incentives have had more to do with their jobs than with the often sporadic performance driven pay systems". Although PRFSs may not be sporadic in that the formulas are applied to every funding allocation, authors seem to agree that prestige is a more important motivating factor, and the influence of independent rankings supports this. In public management, contestability, that is, the prospect of competition (Kettl, 2000, p. 40), is seen as central to the benefits derived from new public management reforms. Universities compete for prestige.

Convergence and challenges

Developing international consensus on PRFS best practice is proceeding slowly because international discussion has been limited. Most in-depth analyses of PRFSs are specific to a nation or even a discipline within a nation. Most cross-national material is not analytical and provides short, static, country-based summaries of complex and evolving systems. The RAE is well known and has been highly influential. All cross-national short

summaries include the RAE, and one suspects all PRFS designers look to the RAE first. Hong Kong, China, copied the RAE. The reverse may not be true. There is no evidence that the UK discussion about revising the RAE is building on lessons learned elsewhere, for example the high level of bibliometric sophistication in Australian or Norwegian efforts. Norway did learn from prior efforts, specifically the RAE and the Composite Index. Denmark is copying Norway. However a host of others are unknown to each other and to the main players. The leadership of international bodies, such as the commissioning of this study and recent reports by the European Union and Hera (Dolan, 2007; European Commission, 2010), is therefore to be welcomed. Consensus and state of the art is much more likely to develop around international rankings of universities. The *Times Higher Education (*or *THE)*, Shanghai Jiao Tong, and emerging corporate products are discussed internationally, are analysed and compared, and are politically influential (Butler, 2010; European Commission, 2010; anonymous, 2010).

Although consensus is slow to develop, the systems exhibit similarities in design because they face similar challenges. Peer review is held in high esteem because it is well accepted by the academic community in every country. However, as it is expensive and time consuming, it is conducted irregularly. As a result, funding decisions may be based on out-of-date information. Departmental level PRFS using peer judgment based on indicators seems to be the state of the art and is being implemented in ERA.

University-level evaluation is metric-based and conducted annually using bibliometric methods. As these are not much liked by the academic community, governments that propose them may face heavy criticism. In smaller nations, the ability to handle very large datasets has made possible departmental-level bibliometrics that are national in scope. However, the departmental affiliation of authors, as indicated on papers, is often ambiguous and cleaning the data is onerous (Debackere and Glänzel, 2004). When funding is allocated on the basis of such data, they must be absolutely clean, because universities would invest much time in challenging their allocations if errors were found. A key advance in the PRFS bibliometric method is the introduction of weighted categories of journals. PRFS frameworks now assign higher weights to the top 10-20% of journals.

The current challenge is to represent adequately the scholarly output of social science and humanities fields. Counting journal articles indexed in databases such as Web of Science or Scopus works for scientific fields but is inadequate for social sciences and humanities, both because the indexing of social science and humanities journals is inadequate and because scholars in these fields produce more than English-language journal articles (Hicks, 2004). For these reasons the Norwegian model, though metrics-based, does not simply count indexed publications but is based on a national research

information system of university submissions of outputs. The Swedish model includes a sophisticated field-based weighting scheme. The ERA suite of indicators varies by field, with scholars in each field choosing the most appropriate metrics for their field. In the Spanish system, researchers are evaluated within one of 11 fields and the criteria used vary somewhat among fields (European Commission, 2010, p. 122; Jiménez-Contreras *et al.*, 2003).

A challenge still unmet in any PRFS is to recognise the broader impact of research, that is, its application and commercialisation. The United Kingdom is currently discussing this. The RQF included measures of broader impact, but was dropped because of its complexity. ERA does not include broader impact measures. The *sexenio* has been criticised for excluding consideration of broader impact (Sanz-Menendez, 1995).

The state-of-the-art in PRFS design incorporates extensive consultation with the academic community. This was not the case initially; the first UK RAE and the Australian Composite Index were imposed. However, in the United Kingdom an extensive public discussion of the RAE ensued, much of it in the pages of the *THE*. Studies of effects were commissioned and the current design process for the RAE's replacement includes extensive formal consultation. Similarly in Australia the design of both the RQF and the ERA involved "exhaustive consultation with researchers and the 39 universities in the Australian system. There is a strong requirement of procedural fairness and transparency and acceptance by key stakeholders." (European Commission, 2010, p. 86) In Denmark years of discussion did not lead to a consensus, thereby preventing the introduction of a PRFS. Lately the discussion has moved forward and the introduction of the Norwegian system is being designed in collaboration with the research community (Fosse Hansen, 2009). In Poland the regulations governing the PRFS are being revised and comments have been solicited from the heads of academic units. The Norwegian model was designed by the ministry in consultation with the university sector, represented by the Norwegian Association of Higher Education Institutions (*i.e.* Rectors' Conference), and this body has ongoing responsibility for the publication indicator (Sivertsen, 2010, p. 2). The increased consultation in PRFS design may signal a shift from the new public management origins of the RAE and the Composite Index to the newer public values/networked governance model in which extensive consultation with stakeholders is preferred (Stoker, 2006).

Notable in PRFSs is a tension between complexity and practicality. The RQF goal of assessing at the level of the research group was too complex to be practical. The complexity of submissions required by the RAE increased over the years, and departments elaborated their submissions over time in an effort to become more competitive. This raised questions about the

cost/benefit ratio of the exercise, and the UK government proposes a metrics-only future in the REF. Complexity emerges in these systems as a response to consultation which produces pressures for fairness across heterogeneous academic disciplines. In Hong Kong, China, extensive consultation prior to the 1999 RAE resulted in a broadening of the activities assessed from research only to include research-related scholarly activities: discovery, integration, application and teaching (French *et al*., 2001, p. 37). Presumably, complexity increases easily in the absence of any accounting of the full cost.

PRFS design and implementation also creates tensions between autonomy and control. This theme is prominent in the English-language commentary accessible to the author. British and Australian commentators on PRFSs are sensitive to the subtleties of rhetoric and reality in relation to university autonomy (Marginson, 1997). Tapper and Salter argue that: "Ironically, such a model of governance may constrain higher education institutions more severely whilst giving the impression, or at least creating the illusion, that university autonomy has been retained" (Tapper and Salter, 2003, p. 11). For example, the Australian Composite Index is used to allocate a block grant; it would seem universities have autonomy in deciding how to spend the block of money they receive. However, this block grant is divided into seven parts, each of which is subject to a separate set of "broad" guidelines by the funding agency which thereby retains a large degree of government control over expenditure. Herbst, an American, observes: "The move of European higher education systems toward managerial autonomy has strings attached: institutional autonomy appears to be offered with one hand – and taken back with the other." (Herbst, 2007, p. 79) In addition, the influence of a PRFS will depend on how universities allocate funding internally; conceivably university management could negate or enhance PRFS incentives (see Chapter 4; also Frölich, 2008). Autonomy, self-governance and competition are sensitive issues for the academic community and each of these is implicated in the introduction of a PRFS.

At the individual level, a great deal of commentary related to autonomy implicitly assumes a prior state in which collegial relations governed the academic community through informal peer review. Others argue that this is an idealised view which neglects professional hierarchies and intellectual authority relations (Harley and Lee, 1997). Whitley develops this argument in expressing his concern about the possible deleterious effects of PRFSs on the vibrancy of the scholarly community and the knowledge it creates. In essence, he argues that strong evaluation systems will reinforce the influence of conservative scientific elders, thereby suppressing novelty, new fields, diversity and pluralism. This problem will be exacerbated if a country's scientific elite is cohesive and if they also control project-based

funding through peer review. The problem will be mitigated if the country has diverse funding agencies with diverse goals, especially public policy and use-oriented goals. University autonomy will also reduce this effect. However, if a system is segmented, that is, applied research is in one place and basic in another, and if career paths are locked into one or the other place, varied funders will not matter because the same elites will control peer review for both the PRFS and the grants associated with universities (Whitley, 2008, pp. 14-16). The problems generated by a cohesive intellectual elite in control of a PRFS may be particularly visible in economics, with its striking division between neoclassical scholars and everybody else (European Commission, 2010, p. 116; Harley and Lee, 1997). This point is connected to the magnified effect of PRFSs through their effects on other parts of the funding system mentioned above. In essence, a PRFS will have contradictory and ambiguous effects on university autonomy, but under the right circumstances a PRFS will certainly enhance control by professional elites.

Conclusion

To conclude, the elements discussed above are drawn together in a consideration of possible futures for PRFSs.

Many countries with PRFSs are embarking on stringent austerity programmes which will prevent increases in funding or lead to decreased funding. In terms of the possible consequences of running a PRFS in an austerity environment, a PRFS generates pressure for increased funding, on the one hand, while on the other, contestability rather than the available funding is the crucial element. Austerity programmes will inevitably lead to dissatisfaction and alienation, which will likely find their way into anti-PRFS rhetoric, but presumably universities can compete as effectively to minimise cuts as they can to maximise funding increases. Thus cuts across the board may not be a good idea in a PRFS system; maintaining consistency and allocating cuts using the PRFS will probably be important.

There is some concern that while the introduction of a PRFS initially brings performance gains, after a few iterations, improvements without funding increases are no longer possible and the costly exercises then return little to no benefit (Geuna and Martin, 2003; Hicks, 2008). The idea is that the PRFS harnesses latent capacity in the system without adding more research resources, but eventually that capacity is exhausted and further gains require more research resources. It would be rash to use this reasoning to remove a PRFS, however, because this analysis neglects the demonstrable benefits of introducing contestability and incentives into a system. The gains in performance are clear when contestability is introduced for the first time,

but eventually improvements slow as a new steady state is reached at a higher performance level. If contestability, which produced the higher-level steady state, were removed, one suspects the higher steady state would not be maintained. Strictly speaking, a government should maintain a PRFS if the difference in research excellence between the before and after steady states is achieved more cost effectively with a PRFS than by adding research resources. Any such calculation should build in the inexorably increasing cost of a PRFS over time as complexity increases as a result of stakeholder consultation. Although possible in theory, in practice it is impossible to make this calculation since the costs of PRFSs are not articulated and the benefits have not been quantified.

Any consideration of the future of PRFSs should take into account relevant non-PRFS mechanisms. Since contestability seems to be at the heart of the benefits of a PRFS, it may be possible to encourage research excellence simply by relying on independent rankings. International rankings are proliferating and their influence is likely to increase greatly over the next few years. Funding could be linked to the results of an international ranking or not. A government could simply rely on media attention and student and faculty pressure to create incentives from rankings. This is in fact a notable element of the highly successful US university system, for which *US News & World Report* annual departmental rankings and National Academy decadal departmental rankings are highly influential. Substituting an international ranking for a PRFS would have the benefit of eliminating costs and shifting divisive and resource-consuming methodological arguments to the international scholarly community. However, universities can only compete effectively for prestige if they have institutional autonomy and discretionary resources (again this is the case in the US system). This recalls the new public management origins of PRFSs. The introduction of a PRFS tends to be just one part of larger changes and although independent rankings may substitute for the research evaluation component, without the larger changes universities cannot respond to incentives to increase their prestige. This suggests that the focus should not be the PRFS *per se*, but rather increasing contestability and institutional autonomy in a university system.

The future of PRFSs also depends on how successful they prove to be in comparison to the alternative centre-of-excellence approach in which governments award a limited number of very large, long-term block grants to universities based on competitive proposals. Germany, Japan and Poland use this approach. China has also taken this approach with its Project 985, though without the competitive proposal aspect. This is clearly a mechanism meant to concentrate funding and encourage international levels of excellence – the same goals that motivate PRFSs. A long-term comparison of the

relative merits of centre-of-excellence and PRFS approaches might help governments understand and explore their options for increasing university research excellence.

The long-term future of a PRFS will depend on how well it meets the government's goals. Many governments have articulated a clear goal for their PRFS: the enhancement of research excellence. A competition for prestige among universities based on research performance will likely achieve that goal. Problems may arise however if governments realise that their goals, or values, are broader. First, a PRFS will not be a good way to encourage interaction with industry and application of research, activities with more demonstrable economic benefits than general research excellence. Enhancing universities' contributions to the economy is a common policy goal that is not well addressed in current PRFSs. Equity and diversity are also important public values in relation to universities. Excellence and equity have always been in tension in research policy. PRFSs encourage excellence at the expense of equity. A more subtle conflict may arise in a PRFS because of the strong reliance on the academic elite in its design and implementation and the possibly enhanced effect if the rest of the funding system is entrained by the PRFS. Whitley suggests that novelty, innovation and intellectual diversity may be suppressed because elites tend to judge academic quality in part on how well work advances paradigms they have themselves established. In addition, the contribution of universities to national and cultural identity may lessen because these are devalued in systems that focus on research excellence at the international level (*i.e.* published in English). There may be circumstances in which any of these value conflicts becomes politically unacceptable. Two choices will present themselves, either further increasing the complexity of the PRFS to broaden the definition of performance (and so increase cost), or reducing the influence of the PRFS and adding another programme to distribute research support based on consideration of other values. The vision of an ever more burdensome PRFS, forced to serve goals for which it is less than ideal seems to hold less appeal than a suite of programmes, each optimally designed to serve a different goal. The second option would increase diversity in the funding system and facilitate differentiation among autonomous universities managed strategically. This should enhance the ability of the system as a whole to serve a complex suite of public values.

As PRFSs are but one element in a complex university governance system, it is likely impossible to devise a single optimal, universal, perpetual best practice. Instability is evident today in the many systems being redesigned. Introducing international discussion in this context will be extremely valuable for comparing options and helping governments choose the best way forward.

References

Anonymous (2010), "The Ratings Game", *Nature*, 464(7285), p. 7, 4 March.

ARC (2009), *ERA 2010 Submission Guidelines*, Australian Research Council, *www.arc.gov.au/era/key_docs10.htm*.

Australian Government Department of Education, Employment and Workplace Relations (2009), *Finance 2008, Financial Reports of Higher Education Providers*, Canberra.

Brookings Institution Press (2005), *The Global Public Management Revolution*, *www.brookings.edu/press/Books/2005/globalpublicmanagementrevolution2.aspx*.

Butler, D. (2010), "University Rankings Smarten Up", *Nature*, 464(7285), pp. 16-17, 4 March.

Butler, L. (2003), "Explaining Australia's Increased Share of ISI Publications – The Effects of a Funding Formula based on Publication Counts", *Research Policy*, 32(1), pp. 143-155.

CHEPS (2007), *Higher Education in Germany: Country Report*, Center for Higher Education Policy Studies (CHEPS), The Netherlands.

Debackere, K. and W. Glänzel (2004), "Using a Bibliometric Approach to Support Research Policy Making: The Case of the Flemish BOF key", *Scientometrics*, 59(2), pp. 253–276.

Department of Trade and Industry (DTI) and Office of Science and Innovation (2007), *PSA Target Metrics 2006*, HMSO, London.

Dolan, C. (2007), *Feasibility Study: the Evaluation and Benchmarking of Humanities Research in Europe*, HERA D4.2.1, Arts and Humanities Research Council, United Kingdom.

European Commission (2010), *Assessing Europe's University-Based Research*, EUR 24187 EN, Science in Society 2008 Capacities, 1.4.1, European Commission, Brussels.

Fosse Hansen, H. (2009), *Research Evaluation: Methods, Practice, and Experience*, Danish Agency for Science, Technology and Innovation, *http://en.fi.dk/publications/2009/research/Research%20Evaluation%20Methods-%20Practice-%20and%20Experience.pdf*.

Franceschet, M. and A. Costantini (2009), "The First Italian Research Assessment Exercise: A Bibliometric Perspective", paper submitted to *Research Policy*, *http://eprintweb.org/S/article/cs/0912.2601*.

French, N.J., W.F. Massy and K. Young (2001), "Research Assessment in Hong Kong", *Higher Education*, 42(1), pp. 35-46.

Frölich, N. (2008), *The Politics of Steering by Numbers: Debating performance-based Funding in Europe*, (No. 3/2008). *NIFU STEP Studier av innovasjon, forskning og utdanning*, NIFU STEP, Oslo.

Geuna, A. and B.R. Martin (2003), "University Research Evaluation and Funding: An International Comparison", *Minerva*, 41(4), pp. 277-304.

Harley, S. and F.S. Lee (1997), "Research Selectivity, Managerialism, and the Academic Labor Process: The Future of NONMAINSTREAM Economics in UK Universities", *Human Relations*, 50(11), p. 1427.

Harman, G. (2000), "Allocating Research Infrastructure Grants in Post-binary Higher Education Systems: British and Australian Approaches", *Journal of Higher Education Policy and Management*, 22(2), pp. 111-126.

Herbst, M. (2007), "Financing Public Universities", *Higher Education Dynamics,* Vol. 18, Springer.

Hicks, D. (2004), "The Four Literatures of Social Science", in *Handbook of Quantitative Science and Technology Studies*, Kluwer Academic Press, Dordrecht.

Hicks, D. (2008), "Evolving Regimes of Multi-University Research Evaluation", *Higher Education*, 57(4), pp. 393-404.

Hicks, D. and J. Wang (2009), *Towards a Bibliometric Database for the Social Sciences and Humanities – A European Scoping Project*, final project report for the European Science Foundation.

Higher Education Funding Council for England (HEFCE) (1997), *The Impact of the 1992 Research Assessment Exercise on Higher Education Institutions in England*, No. M6/97, Higher Education Funding Council for England, Bristol, *www.hefce.ac.uk/pubs/hefce/1997/m6_97.htm*.

Higher Education Funding Council for England (HEFCE) (2006), *Funding Higher Education in England: How HEFCE Allocates its Funds*, Higher Education Funding Council for England, Bristol.

Higher Education Funding Council for England (HEFCE) (2010), "HEFCE: Research: Research Excellence Framework (REF)", *www.hefce.ac.uk/Research/ref/*.

Hodder, A.P.W. and C. Hodder (2010), "Research Culture and New Zealand's Performance-based Research Fund: Some Insights from Bibliographic Compilations of Research Outputs", *Scientometrics*, 84, pp. 887-901.

Hong Kong, China, University Grants Committee (UGC) (2006), *Research Assessment Exercise 2006 Guidance Notes*, *www.ugc.edu.hk/eng/ugc/publication/prog/rae/rae.htm*.

Jiménez-Contreras, E., F. Anegon and E. Lopez-Cozar (2003), "The Evolution of Research Activity in Spain – The Impact of the National Commission for the Evaluation of Research Activity (CNEAI)", *Research Policy*, 32(1), pp. 123-142.

Kettl, D.F. (2000), *The Global Public Management Revolution: A Report on the Transference of Governance* (2nd ed.), Brookings Institution Press, Washington, DC.

Lipsett, A. (2007), "RAE Selection Gets Brutal", *Times Higher Education Supplement* (1779), 1.

Luwel, M. (2010), "Performance-based R&D Institutional Funding in Flemish Universities", presented at OECD workshop, 21 June, Paris.

Marginson, S. (1997), "Steering from a Distance: Power Relations in Australian Higher Education", *Higher Education*, 34(1), pp. 63–80.

New Zealand, Tertiary Education Commission (2010), "Performance-Based Research Fund – Tertiary Education Commission", *www.tec.govt.nz/Funding/Fund-finder/Performance-Based-Research-Fund-PBRF-/*.

Pain, E. (2010), "Polish Science Reforms Bring Fear and Hope", *Science*, 327, p. 1442.

Poland, Ministry of Science and Higher Education (2010), "General Principles of Parametric Evaluation of Scientific Institutions" (in Polish).

Rodríguez-Navarro, A. (2009), "Sound Research, Unimportant Discoveries: Research, Universities and Formal Evaluation of Research in Spain", *Journal of the American Society for Information Science and Technology*, 60(9), pp. 1845-1858.

Sanz-Menendez, L. (1995), "Research Actors and the State: Research Evaluation and Evaluation of Science and Technology Policies in Spain", *Research Evaluation*, 5(1), pp. 79-88.

Sastry, T. and B. Bekhradnia (2006), *Using Metrics to Allocate Research Funds: A Short Evaluation of Alternatives to the Research Assessment Exercise*, Higher Education Policy Institute, Oxford.

Sivertsen, G. (2009), "A Bibliometric Funding Model based on a National Research Information System", presented at the ISSI 2009, Rio de Janeiro, Brazil, *www.issi2009.org/agendas/issiprogram/activity.php?lang=en&id=108*.

Sivertsen, G. (2010), "A Performance Indicator Based on Complete Data", *ISSI Newsletter*, Vol. 6(1), pp. 22-28.

Stoker, G. (2006), "Public Value Management: A New Narrative for Networked Governance?", *The American Review of Public Administration*, 36(1), 41-57.

Strehl, F. (2007), "Funding systems and their effects on higher education systems - international report", *Education Working Paper*, No. 6, OECD, Paris.

Tapper, T. and B. Salter (2003), Interpreting the Process of Change in Higher Education: The Case of the Research Assessment Exercises, *Higher Education Quarterly*, 57(1), pp. 4-23.

Whitley, R. (2008), "Changing Governance of the Public Sciences", in R. Whitley and J. Gläser (eds.), *The Changing Governance of the Sciences*, Sociology of the Sciences Yearbook, Springer, Dordrecht.

Chapter 2

Performance indicators used in performance-based research funding systems

by

Hanne Foss Hansen
University of Copenhagen

This chapter focuses on the different types of indicators used in performance-based research funding systems. The indicators are classified into three groups: first-order indicators, related to inputs, processes, structures, outputs and effects; second-order indicators, which tend to be summary indexes; and third-order indicators, which involve peer review. It analyses their strengths and weaknesses and how they are used in national systems.

Introduction

In recent years more governments have developed performance-based funding systems in tertiary education for both education and research activities (Geuna and Martin, 2003; Whitley and Gläser, 2007; Frölich, 2008). This chapter analyses and discusses performance indicators used in performance-based research funding systems (PRFS) introduced by governments to allocate funds for public research to tertiary education institutions.

Following this introduction the chapter first gives an overview of the variety of indicators. Starting with a brief overall presentation of performance indicators as discussed in public management literature, the discussion explores different types of research indicators such as ratings by peer review panels, indicators reflecting institutions' ability to attract external funding, as well as results indicators (*e.g.* numbers of publications, citations and patents). The following section turns to an analysis of how national performance-based funding systems are constructed, *e.g.* which indicators are used, how they are weighted, which data sources are used and whether systems differentiate their use of indicators across fields, etc. There follows a discussion of consistency in performance measurement, *e.g.* how the use of quantitative indicators compares to peer review processes in terms of capturing performance. The chapter concludes by pointing out knowledge gaps for which further analysis could usefully be carried out.

The focus in the chapter is on indicators used in funding systems based on *ex post* evaluation. Foresight methods and other strategies for identifying knowledge requirements are not explored. Also examined are government funding formulas for institutions. This has two implications. First, that funding systems based on individual contracts between governmental agencies and institutions are not discussed. Second, that governmental project and programme funding as well as research council project and programme funding are not examined, nor are funding formulas used within institutions. Further, as the focus is on funding systems, national research evaluation systems not directly linked to funding are not considered.

Analysing the use of indicators in performance-based research funding systems is comparable to aiming at a fast-moving target. Systems are continuously being redesigned. It is very important to be aware of which system versions are being discussed.

The concept of performance and the rich world of indicators

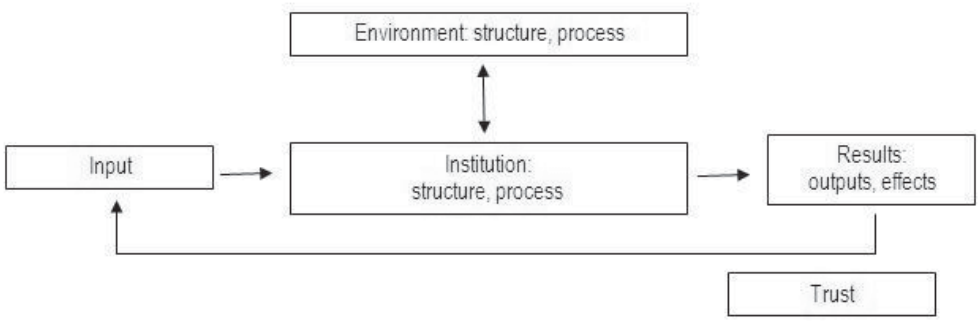
Performance-based management has become widespread in the public and not-for-profit sectors. A variety of terms are used in the literature on this topic. Besides performance-based management there is results management, result-based management, managing for results, managing for outcomes, outcome-focused management and performance management. The intent behind this movement is to measure what results are brought about by institutions and to use that information to help better manage public funds and better report on the use of those funds (Mayne, 2010).

The concept of performance is not unambiguous. Performance must be viewed as information about achievements of varying significance to different stakeholders (Bouckaert and Halligan, 2008). In PRFSs the focus is on organisational performance. Performance is viewed as activity and results and as creating societal value through knowledge production. The underlying idea of PRFSs is that tertiary education institutions may lose sight of the intended results if they are not held accountable by coupling performance to resource allocation (Talbot, 2007). Put another way, PRFSs constitute incentives that can improve performance. In addition PRFSs are anchored in a belief in the possibility of defining and measuring research (and to some extent also research-linked) performance. As will appear, this is difficult in practice. First, there are many different indicators. Second, indicators are proxies and in many respects knowledge about their reliability as proxies is inadequate.

Measuring performance

How can performance – and especially research performance – be conceptualised? Figure 2.1 presents a simple systemic framework illustrating the complexity of organisational performance.

Figure 2.1. A systemic framework for organisational performance



According to the systemic framework illustrated in Figure 2.1, tertiary education institutions are viewed as systems of knowledge production that transform input (funding and other resources) into output (*e.g.* publications and PhD graduates). Knowledge production processes take place in an organisational context constituted by intra-organisational structures (*e.g.* research groups and departments) and inter-organisational environmental structures (*e.g.* disciplinary and interdisciplinary networks).

There is an ongoing discussion in the academic community about the definition of quality in research outputs. Quality is a multi-dimensional phenomenon and includes aspects such as originality, solidity and informativity (*e.g.* Hemlin, 1991; Seglen, 2009). Further outputs are not an end in themselves. Results in terms of effect (sometimes also termed outcome or impact) and benefits are crucial goals, and there are many stakeholders. One is the scholarly community and the contribution research makes to the advancement of scientific-scholarly knowledge. Other groups in society are concerned with the contribution research makes to educational, economical, environmental, cultural and other dimensions of societal development. Research contributions are not always received positively. New knowledge may be critical, provocative or even (considered) harmful. The concept of research performance is not only multi-dimensional and ambiguous. It is may also be charged with conflict.

There are gaps between the output and its effects. First is a gap in time, for example, from knowledge produced to knowledge published and knowledge used. Second, there are discontinuities. For example knowledge may be produced but not published or published but not used. And there is probably an even greater gap between these dimensions and citizens' trust in research institutions. Trust is necessary to maintain public funding in the long run. PRFSs not only aim at creating incentives for and ensuring productivity and effectiveness. They may also play an accountability role with a view to ensuring trust.

In PRFSs research performance is measured by indicators. There are three main categories of research indicators: *i)* first-order indicators aimed directly at measuring research performance by focusing on measuring input, processes, structure and/or results; *ii)* second-order indicators that summarise indexes aimed at providing simple measures of effect (*e.g.* journal impact factor and the H index); and *iii)* third-order indicators from the rating of departments, for example, by peer review panels.

First and second-order indicators, also referred to as metrics, may be used directly and mechanically in funding systems. They may also be used as part of the input to peer review processes, and thus be input to the production of third-order indicators. The following is a discussion of the

content, potential and limitations on the three categories of indicators. In relation to the rich world of first-order indicators, details about the indicators are presented in tables.

First-order indicators

One way to make a typology of first-order research indicators is to distinguish between indicators concerning input, process, structure and results, which again can be divided into output and effects. In the following, important examples of such indicators are discussed. Primary sources are Cave *et al.* (1991), Hansen and Jørgensen (1995), Dolan (2007), Hansen (2009) and European Commission (2010).

Table 2.1. Input indicators

Input indicators	Potential	Limitations
External funding	Ability to attract external funding may be measured as the amount of external research income, perhaps income per full-time equivalent research active staff, and/or as the number and percentage of competitive grants won from selected sources (peer-reviewed external research income, international versus national funding, grants from government including research councils versus grants from industry or foundations). External funding indicators show something about institutions' competitiveness on funding markets and, when defined as peer-reviewed external funding, include an aspect of quality.	Competitiveness does not fully coincide with quality. Reputation and networks also play a role. In addition levels of external funding vary greatly across scientific fields, disciplines and research areas. Differences in levels of external funding combined with differences across institutions in profiles, e.g. whether institutions have a large medical faculty with good possibilities of attracting external funding or a large faculty of arts with less good possibilities, severely limit possibilities for fair cross-institutional comparison.
Recruitment of PhD students and academic staff	Indicators related to the ability to attract students and staff show something about institutions' competitiveness on labour markets and about graduates' and applicants' assessment of the attractiveness of the research environment. Depending on purpose, the number or share of highly qualified and/or international candidates can be counted.	Patterns for applying for university posts are influenced by many factors other than how attractive research institutions are considered in the scholarly community. Research institutions compete not only with other research institutions but also with other career paths.

Important input indicators are presented in Table 2.1. Indicators related to research institutions' ability to attract input in the form of funding from the environment provide information about their competitiveness on funding markets. If measured as the ability to attract peer-reviewed external funding, they also provide information about institutions' reputation and performance

and about the scientific quality and relevance of their research plans. Input indicators do not however fully coincide with quality and performance. Networks for example play a role. Indicators of external funding provide limited possibilities for comparisons across scientific fields because levels of external funding differ across fields.

Indicators related to the ability to attract PhD students and staff provide information about institutions' competitiveness on labour markets as well as about graduates' and applicants' assessment of the attractiveness of their research environment. Application patterns are, however, also influenced by factors such as the attractiveness of other career paths.

Important process indicators are presented in Table 2.2. Process indicators may focus on in-house activities such as the number of arranged seminars and conferences as well as visiting international distinguished academic guests or out-of-house activities such as participation in conferences abroad, invited keynotes and other lectures.

Table 2.2. Process indicators

Process indicators	Potential	Limitations
Seminar and conference activity	The number of arranged seminars and conferences as well as the number of participations in external conferences can be indicators of research intensity.	Conference activity may reflect research tourism.
Invited keynotes	Counting the number of invited keynote addresses given at national and international conferences may be used as a proxy for quality, impact and peer esteem.	Invited keynote addresses may reflect networks rather than quality. No agreed equivalences apply internationally. No possibilities to compare across disciplines.
International visiting research appointments	Counting the number of visiting appointments may be used as a proxy for peer esteem.	Visiting appointments may reflect networks rather than peer esteem. No agreed equivalences apply internationally. No possibilities to compare across disciplines.

Important structure indicators are presented in Table 2.3. Structure indicators may focus on internal aspects such as the share of academics active in research or research infrastructure or on external aspects such as collaborations and partnerships, reputation and esteem. Possibilities for comparisons across disciplines and fields are limited owing to differences in facilities and collaboration patterns.

Table 2.3. Structure indicators

Structure indicators	Potential	Limitations
Staff active in research	Size of active research staff is often regarded as an indicator of research capability and intensity. A sophisticated indicator may measure shares of staff highly active in research, *e.g.* by setting threshold levels of performance for a specific period and counting the number of academics at different levels.	No clear definitions of threshold levels.
Number of PhD students	Number of PhD students is also an indicator of research capability and intensity as active research cultures attract students. Can be measured as mean number per full-time equivalent active research staff.	Disciplines may give different priorities to PhD activity.
Research collaborations and partnership	As research is increasingly conducted in collaborative teams a count of national and international collaborations with other research institutions can be an indicator of research involvement and scale of activity. Research collaborations may be assessed by the degree to which they result in different types of co-publication (national, international, interdisciplinary).	Collaboration is many things. Collaboration may be loose or intensive and mutually binding. Collaboration may involve different institutions, *e.g.* university-university, university-external stakeholder. Collaboration and publication patterns differ across fields.
Reputation and esteem	Positions as journal editors, membership of editorial boards and scientific committees and membership in learned academies are often regarded as indicators of the extent to which researchers are highly regarded by the academic community.	May reflect networks rather than recognition. No agreed equivalences apply internationally. No possibilities to compare across disciplines.
Research infrastructure and facilities	Research laboratory facilities, library facilities (books and electronic journal access), computing facilities, support staff, general working conditions, etc.	Many indicators in one. No easy access to valid comparable data.

Important results indicators concerning output indicators are presented in Table 2.4 and concerning effect indicators in Table 2.5. Output may be measured by counting publications, non-bibliographical outputs, PhD graduates and different kinds of public outreach. Publishing is vital for progress in research but publication patterns differ across fields. Normally publications are counted in groups and most often the focus is on peer-reviewed publications as these are viewed as ensuring a level of quality. In fields such as the social sciences and the humanities, a distinction between the shares of national and international publications may be relevant. Often journal articles are counted in groups according to the ranking of the

journals. In some fields there is high agreement on journal rankings but in others only limited agreement. Effect may be measured by indicators related to citation counts, awards, employability of PhD graduates, commercialisation activities as well as end-user esteem.

Citation counts is an effect indicator in that it indicates how often publications and therefore how often researchers are cited by other researchers. Since researchers cite each other for a variety of reasons, it is, however, debatable what citation counts actually measure. Applicability, visibility and impact are central aspects. To the extent that researchers cite each other because they are building on other researchers' ideas and results, there is a qualitative dimension to citation counts. But how large is this quality dimension? Citation behaviour can also be argumentative (*i.e.* selective as support for the researcher's own viewpoint). It can be used to disagree or to flatter, just as it can be based on a desire to show insight into a subject area (Seglen, 1994a). The intention behind citation counts is often to measure quality, but the information derived relates more to communication structures and professional networks.

Table 2.4. Output indicators

Result indicators: output	Potential	Limitations
Publications	Publishing is vital for progress in research. If counted per full-time equivalent academic staff cross-institutional comparison may be possible. Depending on the purpose, certain types of publication can be counted, *e.g.* percentage of journal articles published in highly ranked journals.	Emphasis on quantity and productivity. Different disciplines produce different types of outputs (journal articles, books, proceedings, etc.). Rating and ranking of journals is not an unambiguous task.
Non-bibliographical outputs	In some fields non-bibliographical outputs such as artworks, music or performances are important.	Due to their heterogeneous character these outputs are not easily measured.
Number of PhD graduates and completion rates for graduates.	New generations of researchers are vital for continuing progress in research. Counts of PhD graduates may be supplemented by a measure of the share of PhD graduates finishing in good time which indicates process effectiveness in PhD programmes.	Disciplines may give different priority to PhD activity and rates of completion may differ across disciplines. Recruitment as well as external employability may affect through-put.
Public outreach	Measures can be developed for the visibility of researchers in society, *e.g.*in the media.	Media visibility may be very loosely coupled to research activities.

Table 2.5. Effect indicators

Result indicators: effect	Potential	Limitations
Citations	Citations provide information about scholarly impact and visibility. Databases such as Web of Science, Scopus and Google Scholar make citation counting possible.	Citations do not fully coincide with research quality. Not all disciplines and research areas are equally well covered in citation indexes. In particular, the humanities and parts of the social sciences and engineering are not well covered.
Number of awards and prizes	Indicator of research quality and impact.	Limited agreed equivalences apply internationally (*e.g.* Nobel prizes). Limited possibilities to compare across disciplines.
Employability of PhD graduates	Industry and governmental employment of PhD graduates can be an indicator of the quality of the graduates and the contribution of research to industry and society.	Employability is sensitive to other factors, *e.g.* regional or national economy. Career paths differ across disciplines.
Knowledge transfer and commercialisation of research-generated intellectual property (IP)	Measure of the extent of income created through patents, licences or start-ups. Important link between IP, commercialisation and economic benefits.	Patents are a poor indicator of IP and sensitive to both discipline and national context.
End-user esteem	Commissioned reports, consultancy and external contracts are measures of the willingness of external stakeholders to pay for and use research. Such measures, *e.g.* counted as the amount and percentage of funding from end users (*e.g.* industry, professions, government, community) are thus indicators of the anticipated contribution of the research.	Different opportunities for different disciplines. Networks influence funding possibilities.

In the academic community there is an interesting ongoing discussion on the use and misuse of citation statistics. In 2008 the Joint IMU/ICIAM/IMS Committee on Quantitative Assessment of Research[1] published a report on the limitations of citation statistics and how better to use them. The background for the report was the observation that the drive for more transparency and accountability in the academic world has created a culture of numbers. The committee wrote: "Unable to measure quality (the ultimate goal), decision-makers replace quality by numbers that they can measure. This trend calls for comment from those who professionally deal with numbers – mathematicians and statisticians." (Adler *et al.*, 2008) Besides the need to consult statisticians when practising citation analyses, the committee noted the scant attention paid to how uncertainty affects analysis and how analysis may reasonably be interpreted. The report has been

commented on by other experts who agree about the common misuse of citation data but also mention ways to make a meaningful analysis, first and foremost by identifying and comparing the performance of comparable groups of scientists (Lehmann *et al.*, 2009).

In addition, there are many challenges as regards technical measurements associated with citation counts. There are competing general databases. The most important are Thompson Reuters ISI Web of Science, Scopus, and Google Scholar. Transparency in the coverage of the databases and the criteria that determine how material is included or excluded leaves something to be desired. Due to the differences in the degree of coverage of the databases, searches on different databases often give very different results. Moreover, the databases are prone to error and the degree of coverage varies from one research area to another. To derive and interpret citation data requires, therefore, a combination of bibliometric skills and specialist scholarly skills in the specific research field. The majority of publication and citation analyses composed by bibliometric experts have used Thompson Reuters ISI Web of Science, the oldest database in this field.

Since both publication and citation patterns vary considerably among research fields, the opportunities for comparison are severely limited. It is therefore recommended only to compare "like with like". For example, to make a citation analysis related to research achievement at a university, the data will have to be normalised, for example, by calculating the average number of citations per article relative to the world average for individual research fields. This makes it possible to show which subjects have more or less impact than would be expected.

In the social sciences and the humanities citation counts have special problems. Journal articles are less important in many disciplines in these fields so that citation analyses produce only partial pictures of performance (for the social sciences, problems and possibilities in citation analysis are thoroughly discussed in Hicks, 2006).

Indicators of knowledge transfer and commercialisation have acquired additional interest in recent years as research and innovation policies have increasingly become integrated. Such indicators are for example concerned with licences and start-ups but may also be related to collaborative research, consultancy activities, networks as well as employability and employer satisfaction with PhD graduates (for an overview see Library House, 2007).

Second-order indicators

As a reaction to the methodological challenges related to citation analyses, several index numbers that are easily accessed in the databases have been developed. It can be tempting to make use of these, but they should be employed with great caution. Unfortunately, this is not always the case.

Two central index numbers are the so-called journal impact factor (JIF) and the H index. JIF is a figure that gives the average number of citations achieved by articles published in a given journal within a given period – the so-called citation window. JIF is a key figure that says something about the journal's characteristics. There are substantial differences in the number of citations to individual articles (Seglen, 1994b). Even in journals with high JIFs, some articles have no or only a few citations. For this reason, JIF should not be used mechanically for the ranking of researchers or research groups, for example.

The H index is defined as the number of articles a researcher has published which have a citation figure equal to or higher than H. An H index of 20 signifies that a researcher has, among his/her publications, 20, each of which has been cited at least 20 times. The H index was developed in recognition of the limitations of other citation measurements. For example, a less significant researcher can have a high total number of citations because he/she has published a "big hit" article with other researchers. To achieve a high H index demands, however, continuous achievement at a high level over a period of years. This also means that the use of the H index only makes sense after 12-15 years of research. In addition, the H index varies according to number of years of employment, the subject and collaboration patterns.[2] The H index therefore does not solve the problem of comparison (Leuwen, 2008). In addition its reliability in general has been questioned and the mean number of citations per paper has been considered a superior indicator (Lehmann *et al.*, 2006).

Third-order indicators

Third-order indicators come from peer review panels that rate, for example, departments. The term "peer review" is used to characterise research evaluation by recognised researchers and experts. Peer review can be described as a collegial or professional evaluation model (Vedung, 1997; Hansen, 2005). The fundamental idea is that members of a profession are trusted to evaluate other members' activity and results on the basis of the profession's quality criteria. Using peer review to produce indicators in PRFSs therefore builds field differentiation into the system even though all fields are treated alike in the evaluation process.

Forms of peer review vary widely (see, for example, Hansen and Borum, 1999; OECD, 2008), and it is useful to distinguish between classic peer review, modified peer review and informed peer review.

Classic peer review is an important mechanism for quality control and resource distribution at the micro level in the research community. Through classic peer review, recognised researchers assess the scientific quality of manuscripts for articles, books and dissertations, and they scrutinise the qualifications of applicants to research posts. Classic peer review is also used in research council systems to determine whether applicants are eligible for support. Classic peer review is linked to clear decision-making situations. Judgements are provided as to whether "products" are worthy of support or publication and as to whether the applicant has the correct qualifications.

Classic peer review that takes place on the basis of the reading of research production is relational in the sense that the assessment is made in a context. A dissertation is assessed, for example, in relation to the research area to which it seeks to contribute, just as an applicant to a research post is assessed in relation to the job description that gives a level and a profile. The process generally includes a form of cross-control of the assessment made. On the one hand, a number of peers may be acting either in parallel or on a panel. On the other, there may be one or more "supreme judges". The assessment of a manuscript for an article is passed on to the editor, who reaches a decision. An assessment of an applicant is passed on for a decision to be made by management.

There is overall agreement that peer review is a reliable method for evaluating scientific quality at the micro level. This does not mean, however, that the method is infallible. There are differences among peers, and there is a degree of uncertainty associated with what in the literature is currently known as "the luck of the reviewer draw". In addition, studies have pointed out that there are biases in some contexts. Bias can be a matter of the "Matthew effect", that is, "to those who have, more shall be given" but bias can also be a matter of systematic unfair treatment or even discrimination on the grounds of gender, age, race or institutional attachment. Networks may make up for discrimination (Wennerås and Wold, 1997).

Over the course of time, other forms of peer review have been developed. After a tentative start in the 1970s, modified peer review has become a commonly used method in some types of PRFSs. As in classic peer review, recognised researchers act as evaluators. But the task and the object of evaluation differ. Modified peer review focuses on the scientific quality of the production of the research organisation. It is most commonly

organised as panel work. The panel members have to cover a larger research area and each individual panel member is therefore a specialist in subsidiary areas of the field to be covered. The basis for assessment most often includes selected publications, but material such as lists of publications, statistics, annual reports and self-evaluations may constitute important background material. When modified peer review is supported by first and second-order indicators it becomes informed peer review.

Summing up on indicators

As has been seen, there is a rich world of research indicators but they are not objective measures of performance. They are proxies, and knowledge about the ambiguity of most of these is limited. For example there is little knowledge about how networks shape measures of institutional competitiveness on external peer-reviewed funding markets and of end-user esteem.

The producers of indicators are creative, and the world of indicators seems to be steadily expanding. There are several reasons for this. One has to do with the ambiguity of indicators. As indicators are proxies with both potential and limitations, their strength is constantly debated. This seems to give rise to ongoing attempts to mend existing indicators to compensate for their weaknesses by developing new ones which have other weaknesses.

Another reason has to do with the development of research policy. At the outset, research policy was built into other policy fields, higher educational policy first and foremost but also sector policies. Then, after the Second World War, research policy increasingly became an independent policy field. Even though research and teaching are still tightly linked at tertiary education institutions, policy streams related to higher education and research are still largely separate and independent. The establishment of independent research policy fields at both international and national levels gave rise to the development of research indicators. In recent years the expansion of research indicators has been furthered by the integration of innovation policy with research policy, which has given rise to new types of indicators related to knowledge transfer and commercialisation.

Because of differences across disciplines and research areas as well as differences in institutional profiles, great care should be taken when using indicators in comparisons. Nevertheless the goal of PRFSs is to make comparisons possible. The following section describes how PRFSs have been constructed and which indicators systems currently in use rely on.

The construction of national PRFSs

As will appear, PRFSs have depended in the past on either third-order indicator models mainly based on peer review or first-order indicator models mainly based on monitoring input and output of research institutions. The analysis of indicators currently in use in PRFSs looks at whether this is still the case or whether they are changing.

The historical background

Historically, two different types of PRFSs have relied on one or the other of these two models. Two countries have pioneered the development of these two types of systems. Table 2.6 provides an overview of the two (for a more thorough comparison, see Hicks, 2009).

Table 2.6. National performance-based funding systems: Historical background

	Third-order indicator model (Britain)	First-order indicator model monitoring input and output (Australia)
Organisation responsible	Higher Education Funding Council for England (HEFCE) among others.	Australian Government: Department for Education, Employment and Workplace Relations
Object of evaluation	Departments (staff actively involved in research hand in publications)	Institutions
Method	Peer review resulting in departmental rating. Peer panel structure as well as rating scales have varied. Rating is subsequently used for distribution of funding.	Indicators used for distribution of funding
Frequency	Exercise conducted 1986, 1989, 1992, 1996, 2001 and 2008.	Annual cycle

The third-order indicator model was developed in England in 1986. The aim was to maintain research excellence by introducing selectivity in funding allocation during an era in which the higher education system was expanding. In 2001 the system was also adopted in Scotland, Wales and Northern Ireland. The British system, called the Research Assessment Exercise (RAE)[3] is based on a large number of peer panels (in 2008 there were 67 panels), each of which assesses and rates the quality of research at all departments in a discipline or given research area. The assessment of the quality of research is based, among other things, on publications by academic staff. The sixth and last assessment round of RAE was conducted in 2008 and will inform research funding in 2009-10. A new system called the Research Excellence Framework (REF) is being developed (see below).

The first-order indicator model monitors input and output and was developed in Australia in 1990. The system, which is still in use while a new system is being developed (see below), monitors four indicators: *i)* institutions' ability to attract research grants in competition (input); *ii)* number of publications (output); *iii)* number of master's and PhD students (stock); and *iv)* number of master's and PhD students finishing on time (output and throughput). The system has been applied uniformly across all research areas using a common list of the types of grants and publications that count (Gläser and Laudel, 2007).

In 2005-07, a second generation, called the Research Quality Framework (RQF), was developed. RQF was a RAE-like system but included in addition assessments by end users of the impact of research on the economy and society. The RQF was controversial as it was considered to lack transparency and had very high implementation costs. When a new government took over in late 2007, the RQF was abandoned prior to implementation. A third generation, called Excellence in Research for Australia (ERA) is being developed and is expected to use first-order indicators as input to third-order peer-review panel assessments.

Both the third-order indicator model and the first-order indicator model have inspired other countries. RAE-like systems have been developed in Hong Kong, China, and in New Zealand and have been proposed but not established in Australia, Denmark, Finland and Sweden. First-order indicator models have been developed in Denmark and Norway.

At present a third model seems to be gaining ground. It could be called a first-order indicator model which monitors, not input and output, but effect. This model is anchored in the idea of counting citations and is being developed in Sweden and the United Kingdom. It is discussed below. The move in this direction seems to be due to a desire to develop "stronger" and more "objective" PRFSs which go beyond outputs in the chain of results to focus on effects. As was noted above, however, it is not obvious that effect indicators are stronger and more objective.

Historically the Flemish Community of Belgium was the first region to experiment with citation counts. In 2003 it replaced a funding formula based on student numbers by a formula called the BOF key which weights student numbers with publications as well as citation counts based on Web of Science data (Debackere and Glänzel, 2004). Over the years, the BOF key has given more weight to publications and citation counts. For 2010 publications and citations are each weighted 17%. The model differs from other country models in that it takes into account differences across disciplines using the journal impact factor.[4]

The Flemish Community has experienced difficulties for applying the system uniformly across all research areas. Subfields of the social sciences

and the humanities, such as law and literature, have proved especially difficult to include. Against this background, the model seems to be shifting towards a more output-oriented model, at least in the social sciences and humanities (see below).

The following discussion further explores the indicators currently in use in PRFSs.

First-order indicator model: monitoring input and output

Table 2.7 gives an overview of countries currently using PRFSs based on first-order indicators mainly based on monitoring input and output.

In 2002, Norway implemented a performance-based funding model combining output indicators (counts of publications and PhD graduates) with an input indicator (external funding).

The publication indicator is based on complete data for the scientific publication output (in journals, series and books) at the level of institutions (Sivertsen, 2006, 2010). The aim of the indicator is to measure and stimulate research activity. The data for the indicator are produced by institutions and included in a national database. The database makes comparable measurement possible as publications are assigned publication points, based on a weighting of publication channels and their quality into two levels, normal level and high level. The latter, which may not account for more than 20% of the world's publications in each field of research, includes leading journals and publishing houses. The list of high-level journals and publishing houses is produced by large groups of peers and is revised annually.

The publication indicator is used for allocating a smaller part of the total direct funding of research in combination with measures of PhD graduates produced and ability to attract external funding. The model is applied uniformly across all areas of research.

The Norwegian model has inspired the development of the Danish model. In Denmark annual increases in resources for block funding of research have for some years been based on a combination of input indicators (external funding and share of educational resources, plus a performance criterion) and output indicators (PhD graduates). From 2010 a publication component is added. The aim of adding the publication component is to encourage researchers to publish in the most acknowledged scientific journals and to strengthen the quality of research. The publication component is similar to the Norwegian publication indicator. A national database is established and publications are divided into publication forms and levels according to lists of journals and publishing houses made by peer groups.

In spite of their similarities, the Norwegian and Danish models are implemented very differently, most probably with different consequences. In Denmark the PRFS is used to allocate the annual increase in block funding for research. The amount of the resources is a decision made by the authorities each year. Part of the increase in resources comes from cutting back existing, mostly historically budgeted, block grants. There is concern in the university sector that increases in resources risk disappearing in the coming years as a result of the economic crisis. If this happens, the PRFS, which currently has marginal importance, may, if it is not re-designed, lose any direct influence.

Table 2.7. PRFSs using first-order indicators monitoring input and output

Country	PRFS	Indicators	Weighting	Data sources	Differentiation
Norway	PRFS was established in 2002 when a partly performance-based research component was introduced as part of the overall funding system. In 2009 the research component distributed 16% of total resources. It is divided into a strategic part (*e.g.* scientific equipment) and a results-based part, called RBO.	RBO has four components: 1) Publications (adjustments for publication form, level and share of authorship); 2) PhD graduates; 3) Ability to attract external funding from the Norwegian Research Council; 4) EU.	The four indicators in the RBO are weighted: - publications 0.3 - PhD graduates 0.3 - Funding (Norwegian Research Council) 0.2 - Funding (EU) 0.2.	National. A national database of publications has been developed.	No differentiation. The economic value of publication counts is equal across fields. Statistics indicate that different areas are treated fairly as they have similar impact in the research component.
Denmark	From 2010 increases in block grants for research at universities have to some extent been allocated across institutions using a PRFS.	Four: - ability to attract external funding; - publications counted (adjustments for publication form, level and share of authorship); - PhD graduates produced; - share of educational resources (which are also allocated on a performance criterion).	From 2012 components are weighted: - external funding 20% - publications 25% - PhD graduates 10% - education share 45%.	National. A national database of publications has been developed.	The publication component is constructed in such a way that it does not alter the relative share of resources between the humanities, social sciences, natural/technical science and medical science. Resources are allocated conservatively across these four fields and then allocated across institutions using publication counts. The economic value of publication counts thus differs across fields.

The Norwegian PRFS reallocates a fixed volume of total block grants annually. As outputs related to publications and PhD graduates have increased, the income per output unit has decreased. In public debate, the issue of whether incentives are reduced over time is raised from time to time. A report from the Norwegian Ministry of Education and Research noted, "As the performance-based reallocation is a competition within a fixed appropriation the reverse side of the coin is a steadily decreasing income per performance unit…. Only in relation to the element related to funding from the Norwegian Research Council has the incentive effect been maintained at a stable level." (Kunnskapsdepartementet, 2010, p. 167; author's translation).

In Denmark the system is constructed so as to not change the share of resources distributed to the humanities, the social sciences, the natural/ technical sciences and the medical sciences. The implication is that the economic value of publication counts differs across fields and may modify the incentive effect of the system. In Norway there is no differentiation across fields but experience so far shows that there is no noticeable reallocation across scientific fields.

The Norwegian model also seems to have inspired Belgium's Flemish Community, which has begun developing a bibliographical database for the social sciences and the humanities. The database includes different types of research outputs, including journal articles, books authored, books edited, chapters in books as well as articles in proceedings. The database is planned to be used for one of the output indicators in the Flemish government's future research funding formula for the universities from 2012, when the BOF key is to be renegotiated.

First-order indicator model: monitoring effect

Table 2.8 gives an overview of the Swedish PRFS in which an effect indicator is an important component.

Table 2.8. PRFS using first-order indicator monitoring effects

Country	PRFS	Indicators	Weighting	Data sources	Differentiation
Sweden	Since 2009	Two: -bibliometric publications and citation counts indicator; -external funding (all external funding sources have equal weight).	Bibliometrics and external funding are equally important indicators.	ISI Web of Science, publication and citation counts are field normalised.	Scientific fields are given different weights which reflect their differences in propensity to score on citations as well as external funding.

In Sweden, a White Paper published in 2007 proposed to develop a performance-based model (SOU, 2007). The aim was to allocate resources according to performance and quality in order to stimulate quality. The proposal was to allocate the total amount of general government university funds across institutions on the basis of quality assessments of research (50%), measures of field-normalised citations (20%), ability to attract external resources (20%), the share of PhDs among staff (5%) and the number of female professors (5%). Since at that time quality assessments were only done in a few tertiary higher education institutions it was proposed in the short run to allocate 50% of the available resources on the basis of the four indicators related to citations, external funding and staff.

In 2009 it was decided to introduce a modified system that allocated resources on the basis of publications and citation counts as well as external funding (Carlsson, 2009; Sandström and Sandström, 2009). The staff elements, including the gender balance, were not included. In the Swedish model, inspired by British plans at the time to replace the RAE with a system producing robust UK-wide indicators of research excellence for all disciplines (see below), the bibliometric indicator based on Web of Science is weighted equally with the measures of external funding. An important aim has been to develop a model that is able to treat all research areas in the same process. In order to meet the challenges of differences among disciplines and of Web of Science coverage, publications and citation counts are field-normalised, and publications in the social sciences and the humanities have considerably more weight than publications in other areas. As a result, the model is extremely complex and quite opaque, except to bibliometrics experts. The model has become so controversial that the Swedish Research Council in 2009 urged its suspension (Vetenskapsrådet, 2009). This has not happened but inquiries and consultations are going on as to how to proceed in the future.

The British experience also shows that developing monitoring effects is not an easy task. In Britain the Higher Education Funding Council (HEFCE) has for some time worked on the development of a second generation of PRFS, called the Research Excellence Framework. Back in 2006 the government announced that a new system should replace the RAE as of 2008. At the time the idea was to produce robust UK-wide indicators of research excellence for all disciplines. The plan was to produce the full set of indicators for the science-based disciplines during 2009 and that these would influence funding allocations from 2010-11. For the arts and social sciences the plan was to phase in the new system gradually while continuing to use peer review (Higher Education Funding Council, 2007).

Observers have characterised this as a move away from the old "subjective" approach to RAEs towards more "objective" methods based on publications and citation counts to gauge quality and impact, plus statistical counts of external research income and postgraduate student activity (Elzinga, 2008). Experienced research policy advisors have expressed scepticism and warned about the myth of "trust in numbers" (Nowotny, 2007).

During the development process HEFCE has asked bibliometric experts at the Centre for Science and Technology Studies (CWTS) at Leiden University for advice concerning measures of citations. A report published in 2008 shows that not only are the arts and social sciences only partially covered in the Web of Science database, parts of technical science and computer science are also not well covered (Moed *et al.*, 2008).

Critical discussions and methodological challenges have forced HEFCE to modify and to some extent roll back plans. The new British PRFS will still be organised with peer panels. The number of panels will however be reduced, as will the number of publications submitted by academic staff. Panels will also make greater use of quantitative indicators, including citation counts where possible. Panels will be asked to rate departments, with a weighting of 60% for research quality, 25% for wider impact of research and 15% for vitality of the research environment. A pilot exercise is currently taking place. Decisions on the configurations of panels and the methods for assessing impact have not been taken. It seems as if Britain is moving towards an informed peer review model with a component based on an effect indicator.

PRFSs currently using third-order indicators

Table 2.9 gives an overview of economies currently using PRFSs mainly based on third-order indicators. The Australian model is not fully implemented but is under development.

Both the Australian and the Polish models are pure informed peer review models. Peers are not required to read publications and rely solely on discipline-appropriate indicators and information. The Australian indicators are planned to capture both research activity and intensity through measures of research income (input), PhD completions and publications (output), research quality through citation analysis (effect, impact) as well as applied research and translation of outcomes.

Table 2.9. PRFSs using third-order indicators

Economy	PRFS	Indicators	Weighting	Data sources	Differentiation
Australia	Excellence in Research for Australia (ERA): Peer panels rely on discipline-appropriate indicators	Discipline-appropriate indicators in four categories:[1] - research quality (ranked outlets, citation analysis, ERA peer review, peer-reviewed research income); - research volume and activity (outputs, income); - research application (research commercialisation income); - recognition (esteem measures).	Not yet decided how to link to funding.	Being developed.	The Australian model differentiates strongly across fields as indicators are discipline-appropriate.
Hong Kong, China	RAE-inspired system	Assessment of quality of recent performance through assessment of active research staff in cost centres.	Not relevant.	Basic research products, primarily publications.	
Poland	Effectiveness indicator for research units	Units are assessed in five categories. Category 1 units have an effectiveness indicator that is more than 30% above the average of the homogenous unit, and category 5 units have less than 70% of the average.	Complex system of weights of many underlying scores.	Annual unit questionnaire on both research and practical applications of research.	Differentiating across 19 categories of homogenous units across three categories of homogenous fields: *1)* humanities, social sciences and arts, *2)* exact and engineering sciences, *3)* life sciences.

1. Australian Government, Australian Research Council (2009), p. 7.

The Polish information collected through questionnaires to research units includes input information (*e.g.* finance), process and structure information (*e.g.* participation in international research projects and infrastructure) and output and effect information (*e.g.* publications, patents and copyrights).

Both models are characterised as third-order indicator models because they have peer panels as their focal point. It may however be debated whether the room for expert opinion is so restricted as to place peers in a primarily administrative role.

In addition to the above-mentioned countries, Spain has a national third-order indicator evaluation system called the *sexenio* because it is performed every six years (Rodriguez-Navarro, 2009). The Spanish system evaluates the research outputs of tenured professors and establishes a salary bonus for each period positively assessed. As this is not a funding system which allocates funds to tertiary institutions, it is not a PRFS according to the OECD definition.

Mixed indicator PRFSs

Italy and New Zealand are countries which mix elements of the different models. Italy decided in 2009 to allocate 7% of block funding to the universities on a performance base. Two-thirds concerned grants for research. Three indicators were used: *i)* peer review ratings carried out in 2001-03 and published in 2006, weighting of 50%; *ii)* ability to attract EU funding, weighting of 30%; and *iii)* share of government competitive grants, weighting of 20%.

New Zealand has had a PRFS since 2001. Three indicators have been used: *i)* peer review inspired by the RAE but assessing research performance of staff rather than departments as such, weighting of 60%; *ii)* number of graduates, weighting of 25%; and *iii)* ability to attract external funds, weighting of 15%. A peer review takes place periodically. One was carried out in 2003 and in 2006 and another is planned for 2012. The two other indicators are measured yearly. The funding period is the calendar year.

Summarising trends in the use of indicators in PRFSs

The analysis of how PRFSs deal with indicators has revealed the following development dynamics and trends.

First, a comparison of the rich world of indicators and the analysis of PRFSs shows that PRFSs use first-order indicators, especially input and results indicators, as well as third-order indicators. Second-order indicators (JIF, H index) are seldom used directly but may be used informally in peer

review processes and may thus influence third-order indicators. The model used in the Flemish Community of Belgium is an exception as it takes JIF into account in trying to correct for differences across disciplines.

Second, among the first-order indicators, input and results indicators are the types mainly used. Process and structure indicators are used in Poland and have been suggested in Sweden but apart from this they are seldom used directly.

Third, within the results indicator types, output (publications) and effect (citation counts) indicators are mainly used. Indicators for outreach, commercialisation and end-user esteem are seldom used. Indicators used in PRFSs overall are mainly what could be termed academic community indicators. Poland uses both academic community and societal indicators and other countries have been discussing the possibilities of including non-academic community indicators, but so far these have been little used. There are probably several reasons for this. Clear non-academic community indicators are not easy to develop and are probably viewed as less legitimate in the academic community.

Fourth, it seems that still more indicators come into use over time. Output indicators for systematically counting publications are developed and effect indicators are increasingly integrated into systems both as a stepping stone for informed peer review and as effect monitoring in the form of citation counts.

Fifth, although over time more indicators are used, the number of indicators often seems to be reduced between the time of discussing how to construct a PRFS to its establishment.

Sixth, third-order indicator systems based on peer review have developed from modified peer review systems to informed peer review systems. This may strengthen systems by making them more transparent and fair. However, this seems to be accompanied by a reduction in the number of peer panels, probably with the consequence of reducing the peer coverage of research fields. It may also make the peer review process more mechanical.

Seventh, as the use of indicators changes, data sources and ways of handling differentiation across fields change as well. This is summarised in Table 2.10.

Table 2.10. Data sources and field differentiation

Model	First-order indicator model monitoring publications	First-order indicator model monitoring citations	Third-order indicator model
Data sources	National databases (self-reporting, validating).	International citation databases (buy-in).	Made up by departments on request in each assessment round.
Differentiation across fields	Handled by peers who group journals and publishing houses in order to produce comparable publication points. Enacted both without field differentiation (*e.g.* in Norway) and with field differentiation (*e.g.* in Denmark).	Necessary as citation counts are not suited for several fields – most of the humanities, several subfields in the social sciences and some in the technical sciences.	Handled by peer panels which translate their qualitative assessment into a rating.

Finally, countries' arguments for introducing PRFSs are very alike overall, and are related to maintaining and promoting excellence and are implicitly or explicitly related to competitiveness compared to other countries. However, their arguments for choices at the model level differ. At this level models are played off against each other, often it seems with rather weak documentation. Arguments at this level are related to the costs of running the systems, the degree of their transparency and fairness, set off against the wish to develop a system that fits all research fields.

At national level political pressures to introduce and maintain PRFSs seems to initiate more micro-level political power struggles in the research environment among actors advocating different models.

Consistency in the measurement of performance?

In the general literature on indicators and performance as well as in the literature on research institutions the characteristics of good indicators are discussed. The argument is that if the indicators do not meet the criteria that define good indicators they are less useful. Table 2.11 summarises three proposals for such criteria.

The proposed criteria have some common features but also some differences. First, what can be termed the methodological strength of indicators has some common features. Good indicators should be relevant, reliable, credible and verifiable. Second, they have to fit the purpose they are used for. They have to be clear, adequate, fit-for-purpose and, especially for research indicators, able to facilitate comparisons. Third, they have to be accepted and trusted, at least to some extent, as well as understandable and fair. Finally there are more "technical" criteria. Good indicators have to be economical, monitorable and available at the right point in time.

Table 2.11. Characteristics of good indicators

Indicators in general: The CREAM test (Kusek and Rist, 2004)	Indicators in general (Mayne, 2010)	Research indicators (European Commission, 2010)
Good indicators are: • Clear (precise and unambiguous) • Relevant (appropriate to the subject at hand) • Economical (available at a reasonable cost) • Adequate (provide a sufficient basis to assess performance) • Monitorable (amenable to independent validation)	Good indicators are: • Relevant • Available (timeliness) • Understandable (clarity, transparency) • Reliable • Credible	Good research indicators are: • Fit for purpose • Verifiable • Fair • Appropriate • Capable of facilitating comparisons across disciplines and institutions

The preceding discussion has shown that these criteria create challenges in relation to the development of PRFSs. Not all indicators are clear. In addition research institutions as well as disciplines and research fields are diverse. No single indicator is capable of capturing their complexity. To do so adequately requires several indicators. But this increases both costs and complexity.

Also, fair comparisons of diverse phenomena are challenging. If effect indicators are used, as for example in the Swedish PRFS, normalising of data and differentiated weighting become necessary. This makes the system very complex, less transparent and harder to understand for persons with limited "technical" skills.

Peer review has been used to deal with diversity and to translate qualitative assessments into ratings that can be used for allocating funding. However, it is both costly and fallible. It is therefore interesting to explore whether this particular indicator is consistent with other indicators. Some studies have contributed to knowledge on this question.

As a follow-up to the 2001 RAE assessment, an analysis of political science was carried out (Butler and McAllister, 2007). The analysis did not include only citations of articles in journals indexed by Web of Science. Instead, it included citations to all publications submitted to the RAE. The analysis showed that the mean number of citations a work attracts significantly improves a department's RAE outcome. This suggested that citations are an important indicator of research quality as judged by peer evaluation. The analysis however also showed that the second important predictor of outcome for a department – slightly less than half as important as citations – was having a member on the RAE panel.

In 2007 the Higher Education Funding Council for England commissioned the Centre for Science and Technology Studies at Leiden University to carry out a study exploring technical issues for developing the REF (Moed *et al*., 2008). As part of the study a more comprehensive analysis was made of the correlation between the 2001 RAE rating and a normalised citation analysis of the papers submitted to the RAE by departments in eight subject groups covering clinical medicine, health sciences, subjects allied to health, biological sciences, physical sciences, engineering and computer science, mathematics as well as social sciences and humanities. Overall the analysis revealed that there seems to be a correlation between citation analysis and peer review, as the normalised citation impact of departments increased with an increase in ratings. It also revealed exceptions. Engineering and computer science departments with RAE rating levels 2, 3a, 3b, 4 and 5 had similar normalised citation impacts; only the citation impact of departments with RAE rating level 5* substantially exceeded that of departments with other ratings. A similar pattern, although with higher impact levels, was found in clinical medicine.

The Italian evaluation has also been followed up by an analysis of the correlation between peer review scores and both article citations and journal impact factors (Franceschet and Costantini, 2009). The conclusions are in line with the above: the higher the peer assessment on a paper, the higher the number of citations that the paper and the publishing journal receive. However, the strength of the correlation varies across disciplines and depends also on the coverage of the discipline in the bibliometric database used. The greater the coverage, the greater the reliability of citation measures. However, there are also examples of papers receiving positive peer judgments but very few, if any, citations, as well as papers obtaining poor peer judgments but significant numbers of citations. It is worth noting that during the peer review process the peers had very limited knowledge about article citations as these were not yet available, although they had access to the journal impact factor. It may thus not be surprising that the analysis revealed a correlation between peer review scores and the JIF.

Knowledge gaps

The lessons learned from the studies mentioned above indicate that there is still a need to look more closely into the correlation between third-order indicators produced by peer review and first- and second-order indicators. More generally there is a knowledge gap in relation to knowledge of how peer reviews are carried out in PRFSs. While there is some knowledge about peer review processes in other types of evaluation systems, such as grant reviews and assessment of interdisciplinary research, there is very limited insight into peer review processes in PRFS contexts.[5]

Overall there is insufficient knowledge about the development process and the dynamics of PRFSs. This chapter, as well as Chapters 1 and 4, shows that PRFSs are spreading rapidly. They are developing across borders as well as developing new generations within borders. It would be interesting to follow up this OECD initiative in the coming years by systematic monitoring and comparative analysis of PRFSs. Such analysis should be carried out by experts at arm's length: they should not be responsible for policy and development of PRFSs.

In the general literature on measuring performance through indicators there is an interesting discussion of the performance paradox, defined as the weak correlation between performance indicators and performance itself (van Thiel and Leeuw, 2002). The performance paradox is reported to be caused by performance indicators' tendency to deteriorate over time as a result of positive learning (performance improves but indicators lose sensitivity to detect bad performance), perverse learning (performance is reported to go up but this is due to manipulated assessments), selection (differences in performance are reduced due to the replacement of poor performers with better performers) and suppression (differences in performance are ignored).

It is beyond the scope of this chapter to discuss the dynamics of these processes in detail. The main point is the existence of a knowledge gap concerning whether and how indicators in PRFSs deteriorate over time. Do PRFS indicators cause positive learning in tertiary education institutions and national research systems or do they cause perverse learning, selection and/or suppression? Do the dynamics differ for different types of PRFSs?

The performance paradox thinking sets a stage for studies of PRFSs which analyse system design, indicators used and their development, destiny or fortune. As mentioned, PRFSs may be seen as systems that constitute incentives to improve research performance. An important question, however, is whether and how these incentives influence the behaviour of academics who are traditionally considered to be motivated by more intrinsic values. Another approach to PRFS studies is to focus analyses on academic staff behaviour and the importance of the context of PRFSs in this respect. Important relevant questions are: Do systems that count publications, such as the Norwegian and the Danish, increase publication performance (more and better publications) or do they result in researchers maximising publication activity through recycling? Do systems that use citations, such as the Swedish, increase research quality or do they advance citation circles?

In short, more knowledge is needed about how contexts and actors' strategies shape PRFSs and about how PRFSs shape actors' strategies and behaviours, as well as about how these dynamics evolve across several generations of PRFSs.

Notes

1. IMU = International Mathematical Union, ICIAM = International Council of Industrial and Applied Mathematics, IMS = Institute of Mathematical Statistics.
2. To compensate for some of these problems other indexes have been proposed. The M index divides the H index by the number of years since the first paper as a way to compensate for junior scientists. The G index is meant to compensate for extraordinarily high citation counts (see Adler *et al.*, 2008).
3. Until 1992 the system was conducted under the heading "research selectivity exercises". The RAE system was conducted for the last time in 2008. A new system based upon a combination of peer review and research indicators is being developed.
4. Information about exactly how this is done does not appear to be readily accessible.
5. An interesting overview on knowledge about peer review is available in Langfeldt (2001) and there is a special issue on peer review of interdisciplinary research in *Research Evaluation*, Vol. 15, No. 1, 2006.

References

Adler, R., J. Ewing and P. Taylor (2008), "Citation Statistics. Joint Committee on Quantitative Assessment of Research", *www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf*.

Australian Government, Australian Research Council (2009), *ERA 2010 Submission Guidelines. Excellence in Research for Australia*, Commonwealth of Australia, Canberra.

Bouckaert, G. and J. Halligan (2008), *Managing Performance. International Comparisons*, Routledge, London.

Butler, L. and I. McAllister (2007), *Metrics of Peer Review? Evaluating the 2001 UK Research Assessments Exercise in Political Science*, Australian National University.

Carlsson, H. (2009), "Allocation of Research Funds Using Bibliometric Indicators – Asset and Challenge to Swedish Higher Education Sector", *InfoTrend*, 64, 4, pp. 82-88.

Cave, M., S. Hanney and M. Kogang (1991), *The Use of Performance Indicators in Higher Education*, London: Higher Education Policy Series 3.

Debackere, K and W. Glänzel (2004), "Using a Bibliometric Approach to Support Research Policy Making: The Case of the Flemish BOF key", *Scientometrics*, Vol. 59, No. 2, pp. 253-276.

Dolan, C. (2007), *Feasibility Study: the Evaluation and Benchmarking of Humanities Research in Europe*, HERA D4.2.1, Arts and Humanities Research Council, United Kingdom.

Elzinga, A. (2008), "Evidence-based Science Policy and the Systematic Miscounting of Performance in the Humanities", paper given at a workshop on evidence-based practice, University of Gothenburg, 19-20 May.

European Commission (2010), *Assessing Europe's University-Based Research. Expert Group on Assessment of University-Based Research*, Brussels.

Franceschet, M. and A. Costantini (2009), "The First Italian Research Assessment Exercise: A Bibliometric Perspective", submitted to *Research Policy*, 14 December.

Frölich, N. (2008), *The Politics of Steering by Numbers. Debating Performance-based Funding in Europe.* NIFU-STEP, Oslo.

Geuna, A. and B.R. Martin (2003), "University Research Evaluation and Funding: An International Comparison", *Minerva*, 41, pp. 277-304.

Gläser, J. and G. Laudel (2007), "Evaluation without Evaluators: The Impact of Funding Formulae on Australian University Research", in R. Whitley and J. Gläser (eds.), *The Changing Governance of the Sciences*, Springer, Dordrecht.

Hansen, H.F. (2005), "Choosing Evaluation Models. A Discussion on Evaluation Design", *Evaluation*, Vol. 11 (4), pp. 447-462.

Hansen, H.F. (2009), *Research Evaluation: Methods, Practice and Experience*, Danish Agency for Science, Technology and Innovation, Copenhagen.

Hansen, H.F. and F. Borum (1999), "The Construction and Standardization of Evaluation: The Case of the Danish University Sector", *Evaluation*, Vol. 5, No. 3.

Hansen, H.F. and B.H. Jørgensen (1995), *Styring af forskning: Kan forskningsindikatorer anvendes?* Samfundslitteratur, Copenhagen.

HEFCE (Higher Education Funding Council for England) (2007), *www.hefce.ac.uk/pubs/circlets/2007/c106_07/.*

Hemlin, S. (1991), *Quality in Science. Researchers' Conceptions and Judgements*, University of Gothenburg, Gothenburg.

Hicks, D. (2006), "The Dangers of Partial Bibliometric Evaluation in the Social Sciences", *Economia Politica*, XXIII, No. 2, pp. 145-162.

Hicks, D. (2009), "Evolving Regimes of Multi-university Research Evaluation", *Higher Education*, 57 (4), pp. 393-404.

Kunnskabsdepartementet (2010), *Tilstandsrapport for UH-institusjoner 2010*, Oslo.

Kusek, J.Z. and R.C. Rist (2004), *Ten Steps to a Result-Based Monitoring and Evaluation System*, The World Bank, Washington, DC.

Langfeldt, L. (2001), *Decision-making in expert panels evaluating research. Constraints, processes and bias*, The Faculty of Social Sciences, Oslo.

Lehmann, S., B.E. Lautrup and A. Jackson (2006), "Measures for Measures", *Nature,* Vol. 444, 21/28 December, pp. 1003-1004.

Lehmann, S., B.E. Lautrup and A. Jackson (2009), "Comment: Citation Statistics", *Statistical Science,* Vol. 24, No. 1, pp. 17-20.

Library House (2007), *Metrics for the Evaluation of Knowledge Transfer Activities at Universities*, Cambridge.

Leuwen, T. van (2008), "Testing the validity of the Hirsh-index for Research Assessment Purposes", *Research Evaluation*, 17 (2), pp. 157-160.

Mayne, J. (2010), "Results Management: Can Results Evidence Gain a Foothold in the Public Sector?", in O. Rieper, F.L. Leuw and T. Ling (eds.), *The Evidence Book. Concepts, generation, and Use of Evidence*, Transaction Publishers, New Brunswick.

Moed, H.F., M.S. Visser and R.S. Buter (2008), *Development of Bibliometric Indicators of Research Quality*, Centre for Science and Technology Studies (CWTS), Leiden University.

Nowotny, H. (2007), "How Many Policy Rooms Are There? Evidence-based and Other Kinds of Science Policies", *Science, Technology and Human Values,* 32 (4), pp. 479-490.

OECD (2008), *The Role of Expert Review in the Evaluation of Science and Technology: Issues and Suggestions for Advanced Practices*, OECD, Paris.

Rodriguez-Navarro, A. (2009), "Sound Research, Unimportant Discoveries: Research, Universities and Formal Evaluation of Research in Spain", *Journal of the American Society for Information Science and Technology*, 60 (9), pp. 1845-1858.

Sandström, U. and E. Sandström (2009), "The Field Factor: Toward a Metric for Academic Institutions", *Research Evaluation*, 18 (3), pp. 243-250.

Seglen, P.O. (1994a), *Siteringer og tidsskrift-impakt som kvalitetsmål for forskning*, Det Norske Radium Hospital, Oslo.

Seglen, P.O. (1994b), "Causal Relationship between Article Citedness and Journal Impact", *Journal of American Society for Information Science*. 45 (1), pp. 1-11.

Seglen, P.O. (2009), "Er tidsskrifts-renommé og artikkeltelling adekvate mål for vitenskapelig kvalitet og kvantitet", in Ø. Østerud (ed.), *Hvordan måle vitenskap?*, Novus Forlag, Oslo, pp. 39-70.

Sivertsen, G. (2006), "A Bibliometric Model for Performance-based Budgeting of Research Institutions", Book of Abstracts, 9th International Science and Technology Indicators Conference, Leuven, Belgium, 7-9 September, pp. 133-135.

Sivertsen, G. (2009), "A Performance Indicator based on Complete Data for the Scientific Publication Output at Research Institutions", *ISSI Newsletter*, Vol. 6(1), pp. 22-28.

SOU (2007), *Resurser för kvalitet,* Utbildningsdepartementet, Stockholm.

Talbot, C. (2007), "Performance Management", in E. Ferlie, L.E. Lynn and C. Pollitt, *The Oxford Handbook of Public Management*, Oxford University Press, Oxford.

Thiel, S. van and F.L. Leeuw (2002), "The Performance Paradox in the Public Sector", *Public Performance & Management Review*, Vol. 25, No. 3, pp. 267-281.

Vedung, E. (2007), *Public Policy and Program Evaluation*, Transaction Publishers, New Brunswick.

Vetenskapsrådet (2009), *Bibliometrisk indicator som underlag för medelsfördelning. Svar på uppdrag enligt regeringsbeslut*, U2009/322/F (2009-01-29) till Vetenskapsrådet. Stockholm, 27 May,.

Wennerås, C. and A. Wold (1997), "Nepotism and Sexism in Peer Review", *Nature*, Vol. 387/22.

Whitley, R. and J. Gläser (eds.) (2007), *The Changing Governance of the Sciences*, Springer, Dordrecht.

Chapter 3

Performance-based funding for public research in tertiary education institutions: Country experiences

by

Sarah Box
OECD Secretariat

This chapter presents the results of a survey on funding models. It looks at rationales for the use of performance-based funding systems and outlines the key features of schemes and systems currently in use, including frequency of assessment and the allocation of funds. It discusses the effects of using performance-based funding as well as interactions with other funding mechanisms.

Introduction

This chapter draws on responses to a questionnaire distributed to OECD countries and select non-member economies to present a sample of country experiences in the use of performance-based funding for research in tertiary education institutions (TEIs). The survey was part of the project on performance-based funding for public research in TEIs of the OECD Working Party on Research Institutions and Human Resources (RIHR). Some governments use such funding to allocate a portion of research funds to TEIs at the institutional level, generally based on *ex post* evaluation of research outputs and outcomes. The questionnaire sought country-level information on the use of performance-based funding, its features and its perceived impacts. It also sought the views of countries not currently using performance-based funding but which had previously considered or were now considering using such systems. Thirteen responses were received: twelve from countries actively using performance-based funding in some form and one from a country considering its use.

This work can be seen in the context of governments' efforts to improve steering and governance within the research environment. Over recent decades, funding arrangements have evolved in several countries to include a larger share of project funding, and current efforts on performance-based funding may be regarded as a complementary policy tool aimed at improving the outcomes from institutional funding streams.

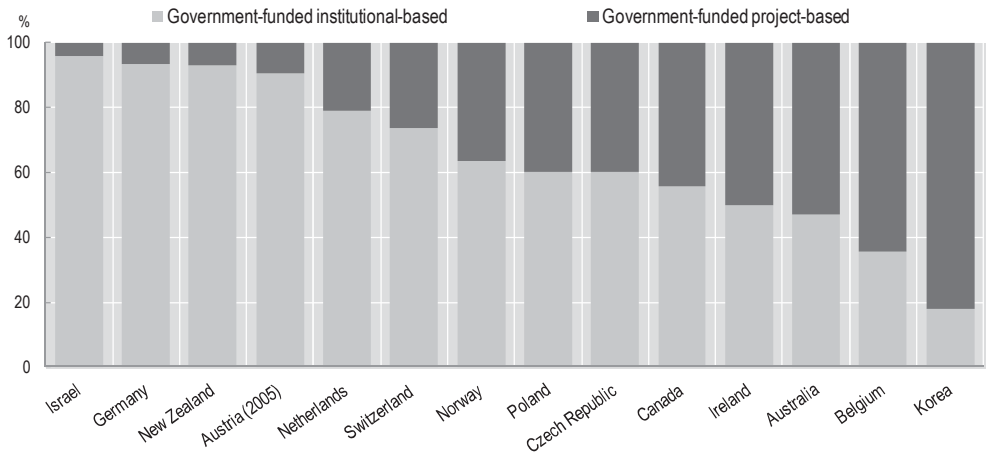
The chapter begins with a brief discussion of general trends in public funding systems for research. It then outlines the findings from the questionnaire, addressing the rationales, key features and effects of performance-based funding systems currently in use. Finally, it summarises key findings and presents a number of policy issues that emerged from the analysis.

Institutional and project funding – the backdrop

Institutional (or block) funding and project funding are the two main mechanisms used by governments for financing research in TEIs. Institutional funding can broadly be defined as funds attributed to universities and research organisations with no direct selection of projects or programmes to be performed, while project funding is attributed to a group or individual to perform an R&D activity that is limited in scope, budget and time and is normally allocated on the basis of a project proposal describing the research to be done (see Lepori *et al.*, 2007a; van Steen, 2010).[1]

Project funding for public research in the aggregate (*i.e.* TEIs plus government research institutions) has grown over time in several countries. For instance, Lepori *et al.* (2007a) found that project funding had increased since 1970 in Austria, France, Italy, the Netherlands, Norway and Switzerland, both in real terms and as a percentage of GDP. They concluded that project funding had also increased as a share of total funding, since total public research funding did not strongly increase over the period. Preliminary analysis from the OECD's National Experts on Science and Technology Indicators (NESTI), reported in van Steen (2010), found a rising share of project funding from 2000 to 2008 in Australia and Austria, although there was relative stability in the modes of public R&D funding in the other countries studied.

Figure 3.1. Government funded R&D in higher education, by type of funding, 2008



Note: This is an experimental indicator. International comparability is currently limited.

Note on Israel: The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Source: OECD, Working Party of National Experts on Science and Technology Indicators (NESTI) project on public R&D funding, 2009.

Nevertheless, despite the general trend towards project funding, some countries still have a high share of institutional funding in their overall funding mix and arrangements vary widely across countries. Austria, Germany, the Netherlands and Switzerland had shares of institutional funding in total national public funding to national R&D performers of over 70%. Australia, Canada, the Czech Republic, Israel and Poland used institutional funding for 50-70% of total funding, while Belgium, Ireland, Korea and New Zealand instead used project funding to allocate more than

50% of funding (van Steen, 2010). These results suggest that the higher education sector is the main destination of institutional funding, although half the countries also provided substantial institutional funding to the government sector. Figure 3.1 shows the proportion of institutional and project funding in the higher education sector in a sample of countries; the share of institutional funding ranges from a high of almost 96% in Israel, to around 18% in Korea. Interestingly, project-based funding shares are much lower in the higher education sector than in public research overall in Israel and New Zealand.

Given the ongoing importance of institutional funding in many countries, coupled with the interest of governments in more effectively steering research activities, it is not surprising that mechanisms that allow some element of direction of institutional funding flows have emerged. Jongbloed (2010) noted that funding is part of the set of governance tools that enforce common goals for higher education, set incentives for certain behaviour and attempt to maximise the desired output with limited resources. Institutional funding generally provides institutions with more scope to shape their own research agenda, while project funding provides governments with more scope to steer research towards certain fields or issues. Project funding may also allow governments to target the best research groups or support structural change (Lepori *et al.*, 2007b). Over recent decades, competitive funding has increased in order to enhance efficiency and quality, and institutional budgets have become more tied to specific teaching and research outcomes via formula-based funding systems (Jongbloed, 2010). In contrast to the 1990s when only a few countries used output-related criteria in funding, almost 20 European countries now use elements of performance to drive budgets of HEIs.

Systems of performance-based funding

Responses to the OECD RIHR questionnaire on performance-based funding for public research in tertiary education institutions were received from Australia, Austria, Belgium, the Czech Republic, Denmark, Finland, Germany, New Zealand, Norway, Poland, Slovenia, Sweden and the United Kingdom. The Belgian response detailed the policies and systems in place in the Flemish Community of Belgium. All references to "Belgium" or "Belgian" thus relate to the Flemish Community. The German response was drawn up on the basis of submissions from nine German regional states (*Länder*): Baden-Württemberg, Bavaria, Berlin, Brandenburg, Hamburg, Hesse, Lower Saxony, North Rhine-Westphalia and Rhineland-Palatinate. In Germany, the *Länder* determine the design and implementation of procedures to fund public higher education, and this has led to a diversity of models for performance-based funding. Therefore, this chapter will refer to both general German conditions and state-specific procedures.

The questionnaire specified that the focus of the survey was the systems of *ex post* evaluation of research outputs and outcomes used to allocate a portion of institutional-level funding for research in TEIs. It noted that the study concerned government-level funding formulas for institutions, not systems for the allocation of project or programme funding to groups or individuals, or the internal allocation of funds within institutions. The following therefore describes funding schemes identified by the surveyed countries as meeting the criteria for inclusion in the study as performance-based funding schemes for public research in TEIs. The interpretation of these parameters by individual countries and the resulting list of schemes discussed in this chapter may not accord with other interpretations and lists (see Chapter 1, for example).

Dates of introduction

The use of performance-based funding systems is generally relatively recent. In the sample of countries responding to the OECD questionnaire, most funding policies have been introduced since 2000, although some schemes have been operating for longer. Of the schemes with longer histories, the "first vintage" of Belgium's main performance-based system for allocating funds for basic research was created in 1985, although the performance-based mechanism with bibliometric indicators was introduced in 2003 and a second fund for strategic and applied research was introduced in 2005. The United Kingdom introduced quality-related (QR) research block funding to institutions in 1989. This system phased out a less structured algorithm based on quality and volume of research. The first Research Assessment Exercise (RAE), currently the United Kingdom's main performance-based funding tool, took place in 1992. In Poland, the performance-based system was introduced in 1991, together with the foundation of the State Committee for Scientific Research, a key institution in the post-1989 arrangements for science and research. Norway began funding part of the total public general university fund on the basis of performance in 1992, and the current system took effect in the 2003 budget.

Some German states introduced forms of performance-based funding allocations for higher education institutions (HEIs) as early as 1994, with some subsequent modifications, while others introduced their schemes after 2000. The "oldest-serving" model, still in use today, is in Rhineland-Palatinate, which has used indicator-based funding for materials and capital expenditure for research and teaching since 1994 and for staff costs since 1998. North Rhine-Westphalia began distributing funding for student assistants, digital literature and equipment on the basis of indicators in 1994, though the current system dates from 2006. In other states, performance-based funding is more recent, dating from 1999 in Bavaria, 2000 in Baden-

Württemberg, 2003 in Hesse, 2004 in Brandenburg, 2004-05 in Hamburg (an earlier system was introduced in 2002-03), and 2006 in Lower Saxony (with an earlier system in place since 2000 for universities of applied science). Berlin is about to introduce a new system of performance-based funding, with grants allocated for the first time using the new criteria in 2012.

Of Australia's current set of schemes funded via performance-based formulas, one (related to research infrastructure) took effect in the 1995 budget, while the others have taken effect since 2002. Australia's new Excellence in Research for Australia (ERA) process for research evaluation, which will be used in funding allocation decisions, begins from June 2010. Finland began allocating part of university resources on the basis of performance in 1998; the current system dates from 2010. The Czech Republic implemented its first model (covering part of TEI research activity) in 2003; a second model was introduced in legislation in 2008 (with budget impacts in fiscal year 2010). New Zealand implemented its Performance-Based Research Fund (PBRF) in 2003, with funding phased in over three years. The first of the periodic quality evaluations took place in 2003, with another partial round in 2006 and a further round scheduled for 2012. Of the most recent schemes, Austria's performance-based system began in 2007, and Sweden's performance-based system took effect in 2009. Denmark's funding via a bibliometric indicator begins in 2010, following a political agreement in 2009. On average, it appeared to take countries two to three years to move from discussions of a particular performance-based funding scheme to the introduction of schemes in budgets.

Rationales for the introduction of performance-based funding systems

Among the rationales for the use of performance-based funding for research in TEIs, many of the questionnaire responses mentioned the search for research quality, the central rationale of several systems. For example, strengthening research quality was the main rationale behind the Danish system, with incentives for publishing in the most recognised journals and by the most recognised publishers. Quality was also the top factor behind the Finnish system, with an additional rationale of demonstrating to the public that research funding is spent optimally. In Norway, the main purpose of the performance-based allocation system is to enhance the quality of research by motivating institutions to increase their research activities and by distributing resources according to research results. This focus accords with the principles set out in the 2000/01 White Paper to the Norwegian Parliament (Storting) *Do Your Duty – Demand Your Rights: Quality Reform of Higher Education*, which was drafted in response to a Royal Commission report on HEIs. Sweden also cited improved quality in research as the main

purpose of the system. The main purpose of the United Kingdom's system is improved research quality, but accountability is also paramount, and making quality visible to the public is an important aspect of this. For the United Kingdom, concentration in research is regarded as a consequence of the system rather than a goal.

However, countries are also seeking to use performance-based funding to create incentives for other actions and behaviours related to research. For instance:

- As well as identifying excellence and improving research quality, Australia's new ERA initiative aims to compare research efforts against international benchmarks and identify emerging research areas for development. More broadly, the funding system is aligned with the principles for higher education research funding in the 1999 Knowledge and Innovation Framework. These principles focus on the achievement of world-class research and research training, encouragement of critical mass in areas of strength, the development of a diverse research base and the ability to respond to global market opportunities, university autonomy in setting research priorities and conducting research, increased collaboration, greater contestability and transparency in allocation processes, and openness to verification.
- Belgium established its performance-based funding mechanism to stimulate scientific performance and the quality of research, to make quality visible, to create an incentive for technology transfer and to distribute research funding on an equitable basis.
- At first, the purposes of the Czech system were to determine where to concentrate research funding, to achieve better organisation, uniformity and transparency in the management of public financial resources for research, and to have effective and optimal spending. Subsequently were added research quality, accentuation of the competitive advantage of Czech research results, and making quality visible to the national and international public.
- The main purpose of New Zealand's PBRF is to ensure that excellent research in the tertiary education sector is encouraged and rewarded, thus increasing quality. It is also designed to achieve other aims, such as improving the quality of public information on research and preventing undue concentration of funding. The design of the scheme also attempts to support Māori and Pacific research capability.

- Poland's performance-based system aims to create incentives to improve research quality and to make this quality visible to the public. The system also seeks to concentrate research funding (or to avoid fragmentation). As part of Poland's shift away from a centralised system of science and higher education, the performance-based funding scheme aims to move decisions on science towards the research community and away from politicians and officials.

Box 3.1. Funding research in Slovenia

In 2010, the Slovenian Ministry of Higher Education, Science and Technology is preparing new strategic plans for higher education and R&D – the National Plan for Higher Education 2011-15 and the National Research and Development Programme 2011-15. In this context, the issues of funding and introducing "novelties" into financial systems are of great relevance and interest.

Slovenia has five universities and 26 other HEIs, of which 20 are also registered as research organisations. Data show that 36% of all researchers work in higher education and more than one-third of research is carried out in HEIs. Currently, public money for R&D activities (basic and applied research) is distributed to public universities and public research institutes exclusively through public calls and tenders, which are based on competitive principles. The main evaluation criteria are quality of research outputs (publications), quality of research programmes proposed (peer review) and connections with industry (money from external sources).

The Slovenian Research Agency (SRA) is responsible for the execution of public research financing and it provides independent decision making on the selection of programmes and projects financed from the state budget and other financial sources. The SRA also evaluates project proposals and the results of science policies. Its evaluation system collects a large amount of information, including data on publications (using the national bibliographic system COBISS and the Web of Science), data on results (using the Slovenian current research information system SICRIS), data on the connection of researchers with the business sector, participation in EU projects, and connection of researchers in the education process. Transparency of criteria, procedures and results is one of the top priorities of the SRA to ensure greater legitimacy of the system and better access to information.

With respect to teaching, the traditional method of funding according to numbers of teachers, workloads and hours per week is gradually being replaced by a funding formula based on numbers of students and graduates by field of study. In 2010, 40% of funding is flexible, with a lump sum being allocated according to the funding formula, while 60% remains fixed. HEIs have full freedom to manage these funds.

Source: Slovenian government response to the OECD RIHR questionnaire on performance-based funding for public research in tertiary education institutions (February 2010).

In contrast to these examples, Austria's indicator-based budget for TEIs does not have a research-specific rationale, as it is used for the institution as a whole and encompasses education, graduates, mobility and gender mainstreaming as well as research. Similarly in Germany, institutional funding does not distinguish between research and teaching and, as such, the objective of performance-based allocations is broader. The general aim is to strengthen the autonomy and institutional responsibility of HEIs, especially in relation to budget management (the implementation of performance-linked models is occurring in the general context of governance reforms with a view to less state intervention, more TEI autonomy and more market activity). The *Länder* also cited the introduction of competition between institutions to encourage optimal use of state grants, clear incentives for improvements in performance and quality, transparency and predictability of budget allocations, and stronger incentives for third-party funding and doctoral graduates.

Given its potential impacts on incentives for quality research and other behaviours, performance-based funding is attracting attention in other countries as a tool for governing and steering research. Slovenia is currently considering the issue of performance-based funding for public research in TEIs in the context of setting broader higher education and R&D development plans (Box 3.1). With experience in providing formula-based lump sum funding for teaching in HEIs, and a wealth of research-related information already gathered by the Slovenian Research Agency, Slovenia would appear to have a good base for introducing performance-based funding for research in TEIs if it chooses to do so.

Key features of systems

More than half of the countries responding to the questionnaire make no distinction between private and public institutions for the purposes of performance-based funding. In Australia, for example, payments are made to eligible higher education providers (HEPs); these are listed in the *Higher Education Support Act 2003*. Eligibility for this list relies on factors related to tuition assurance, accreditation, and quality and accountability; both public and private institutions may be eligible. The Belgian and Finnish systems cover both private and public institutions.[2] Following a legislative amendment in 2008, the Czech system covers all types of research organisations (not just TEIs) if their primary purpose is to carry out and disseminate research, reinvest profits, improve research capacity or results and foster co-operation between the academic and private sector. New Zealand's scheme is open to all New Zealand-based degree-granting tertiary education providers (and their wholly owned subsidiaries) and thus encompasses universities, institutes of technology, polytechnics, *wānanga*[3] and private training establishments.

Norway's funding system covers both private and public institutions, as does that of Poland and Sweden.

In contrast, the performance-based funding systems in Austria, Denmark, Germany and the United Kingdom focus on public institutions. In particular, Austria's system covers its 21 public universities, and Denmark's its eight universities. Germany's performance-based grant allocation procedures encompass public universities and universities of applied science (*Fachhochschulen*) in all *Länder*, as well as public colleges of art (*Kunsthochschulen*) in Berlin, Brandenburg, Hamburg and Hesse, and university clinics in Bavaria, Hamburg and North Rhine-Westphalia. In addition, Baden-Württemberg includes teacher training colleges (*Pädagogische Hochschulen*) in its system, and Bavaria also includes university libraries. All HEIs in the United Kingdom are eligible to submit their research to the Research Assessment Exercise, if they choose to do so, with HEIs defined within relevant higher education legislation. However, private institutions are not eligible.

Table 3.1 provides an overview of the schemes and systems of performance-based funding in the responding countries. The functional detail of the assessment systems can be complex and more details, where available, can be found on the project website.[4] In all countries, the schemes cover all types of research and all fields. Most of the schemes use annual assessment processes, although Austria has a three-year cycle in accordance with its budget, and New Zealand, Poland and the United Kingdom have aspects of assessments taking place every five to seven years. To put performance-based funding into practice, most countries use a "master" performance assessment system that affects a tranche of block funding. However, Australia and Belgium use various performance-based schemes at the level of individual funding programmes. The indicators used by countries are generally quite similar, drawing mainly on information about publications, research income and students. However, Australia's ERA scheme intends to look also at indicators of application/commercialisation and recognition, Belgium uses information on patents and spin-offs, Finland accounts for teacher and researcher mobility, and Poland looks at the creation of new technologies and patents. In contrast to indicator-led models, the United Kingdom's RAE uses peer review techniques, whereby panels of experts assess the research submitted by institutions, while New Zealand's PBRF uses a mixture of periodic peer review and annual quantitative indicator assessments. Poland's system also incorporates some aspects of peer review and subjective decision making. In Germany, peer review and bibliometrics play almost no role in the *Länders'* performance-based systems.

Table 3.1. Key features of performance-based funding schemes in operation

Country	Frequency of assessment	Funding tool	Indicators used	Budget affected
Australia	Grants provided on annual basis Indicators often look at performance in the two most recent years	Research block grants to eligible higher education providers, including: * Joint Research Engagement (JRE) initiative (formerly the Institutional Grants Scheme – IGS) * Research Infrastructure Block Grants (RIBG) * Sustainable Research Excellence in Universities (SRE)	Indicators include: 1. Research income received by HEIs (averaged over two most recent years of data) 2. Institutions' capacity to disseminate research results in (mainly) peer-reviewed research publications (averaged over two most recent years of data); the four categories are books (weighted by factor of 5), book chapters, journal articles and conference papers 3. Institutions' Commonwealth funded higher degree by research (HDR) student load[1] (weighted by course cost) Another indicator, used as a performance moderator, is number of research active staff (averaged over two most recent years of data); research active staff are defined as research only and teaching and research staff. The new Excellence in Research for Australia (ERA) system will use indicators of: a) Research quality: ranked outlets, citation analysis, ERA peer review, and peer-reviewed Australian and international research income b) Research volume and activity: total research outputs, research income and other research items within the context of the profile of eligible researchers c) Research application: research commercialisation income and other applied measures d) Recognition: a range of esteem measures	Funding pools are fixed and annual grants are based on each university's performance relative to others In 2010, AUD 1.42 billion will be provided as block grants to support research and research training in Australian higher education providers through performance-based schemes The JRE incorporates a "safety net" feature that ensures institutions' grants do not fall below 95% of the previous year

Table 3.1. Key features of performance-based funding schemes in operation *(continued)*

Country	Frequency of assessment	Funding tool	Indicators used	Budget affected
Australia *(cont'd)*		Performance-based block funding to support training for students undertaking doctorate and masters degrees by research, including: * Research Training Scheme * Commercialisation Training Scheme * Australian Postgraduate Awards * International Postgraduate Research Scholarships	1. Successful completions of HDR degrees by students in HEIs (weighted by level of course and course cost; data averaged over two most recent years)	
Austria	Every three years (in line with the tri-annual budget system for public universities)	Direct public funding of universities (GUF)	The major R&D indicators are: 1. Number of PhD graduates grouped by field of study 2. Level of external funding for requesting research (*e.g.* local research funds, EU) 3. Level of contract research The indicator-based budget for public universities in Austria is used for the institution as a whole, and thus also includes weighted indicators for the areas of education and social objectives	20% of the direct public funding of universities (GUF)

Table 3.1. Key features of performance-based funding schemes in operation *(continued)*

Country	Frequency of assessment	Funding tool	Indicators used	Budget affected
Belgium – Flemish Community	Annual	Funding to universities via: * BOF (Special Research Funds) * OF (Industrial Research Funds) * Research component of operational remittance	Indicators are calculated for individual institutions as the proportion of the total achieved by all institutions and include: 1. Bachelor and initial masters diplomas 2. Doctorates 3. Annual operational remittance/number of scientific personnel 4. Publications and citations 5. Industrial contract income 6. Income from the European Framework Programme 7. Patents 8. Spin-offs A mobility and diversity parameter is also used	In 2008, 75% of the research funding paid at the institutional level by the Flemish government was subject to performance-based funding via the BOF, IOF or operational remittance for a total of around EUR 312 million
Czech Republic	Annual	Performance-based targeted support for specific university research; and performance-based institutional support	For specific university research funding (student research) (Model 1): 1. Bibliometric indicators 2. Number of PhD students in accredited study programmes 3. Number of magister (master) degrees awarded in last academic year 4. Number of PhD degrees awarded in last academic year For institutional funding (Model 2): 1. Bibliometric indicators	In 2009, CZK 1 047 million was allocated for performance-based university research (this equated to around 4% of total government R&D expenditure and 7% of public general institutional university funds). In 2010, CZK 937 million is allocated for specific university research (almost 4% of total government R&D expenditure). One-third of total government R&D expenditure will be spent on performance-based funding for public research in research organisations including TEIs.

Table 3.1. Key features of performance-based funding schemes in operation *(continued)*

Country	Frequency of assessment	Funding tool	Indicators used	Budget affected
Denmark	Annual	Restructuring fund (a fund additional to the general university fund)	Indicators include: 1. Share of educational resources 2. External research funding obtained by universities 3. Number of PhD graduates 4. The Danish Bibliometric Research Indicator	In 2010, 10% of the restructuring fund was allocated according to bibliometric performance. This tranche equated to DKK 30 million (approximately EUR 4 million). Institutions received a share equivalent to their share of publication output.
Finland	Annual (based on three-year averages in most of the indicators)	Core funding	Indicators include: 1. Teaching and research person-years 2. Number of doctoral degrees determined in the agreement between the Ministry and the institution 3. Number of doctoral degrees completed 4. Nationally competed research funding (Academy of Finland funding, funding via decisions on Centres of Excellence, Tekes funding) 5. Scientific publications (refereed international publications and other scientific publications) 6. Internationalisation of research (amount of internationally competitive research funding and extent of teacher and researcher mobility)	Universities receive a share of criteria-based funding. Around 34% of core funded is performance based.

Table 3.1. Key features of performance-based funding schemes in operation *(continued)*

Country	Frequency of assessment	Funding tool	Indicators used	Budget affected
Germany	Once every financial term (either each year, or in some states for two years)	State-level allocations of institutional funding	Indicators vary by state and by type of institution. Most frequently used are indicators of third-party funding and number of completed doctorates. Some states also use collaboration-related indicators, while Bavaria and Berlin make use of publications data.	Varies by state, with Berlin having the highest share (30%) of performance-based funding for research (from 2012). Approximate shares in other states: North Rhine-Westphalia 11%; Baden-Württemberg and Hesse 10%; Brandenburg 8%. Most states limit the maximum possible budget loss of HEIs through application of a tolerance band based on the previous year's budget
New Zealand	Annual funding allocation but mixture of periodic and annual assessment of performance	Performance-Based Research Fund (PBRF)	Three indicators are used to determine funding allocations: 1. Results of periodic quality evaluation (a peer review process) 2. Number of research degree completions (annual) 3. External research income (annual)	Institutions receive allocations based on relative performance, with yearly changes based on movement in research degree completions and external research income. In 2008, PBRF funding accounted for NZD 231 million (11.8%) of government funding to tertiary education institutions. The budget is set annually by the government.

Table 3.1. Key features of performance-based funding schemes in operation *(continued)*

Country	Frequency of assessment	Funding tool	Indicators used	Budget affected
Norway	Annual (based on results two years earlier)	Performance-Based Reallocation (PBR) of total funding	Four indicators determine the PBR: 1. Publication points 2. Funds from the EU Framework Programme for research 3. Funds from the Research Council of Norway 4. Number of doctoral degrees awarded	Each year, 30% of total funding to the tertiary education sector is distributed among institutions on the basis of relative performance in education and research. In 2010, approximately 6% of total funding to tertiary education institutions was distributed through the PBR for research.
Poland	Five-year assessment of performance	Institutional funding	Indicators include: 1. Reviewed publications 2. Scientific monographs 3. International research projects 4. Authorisation for granting research degrees 5. New technologies, materials, products, systems, services, methods and software 6. Patents and utility models granted, copyrights	In 2008, 60% of GBAORD was allocated through institutional funding, which is partially performance-based
Sweden	Annual	Direct funding	Indicators used are external funding (50%) and bibliometric indicators (50%)	All new direct funding since 2009, plus 10% of every institution's direct funding from the previous year is allocated according to the performance system

Table 3.1. Key features of performance-based funding schemes in operation *(continued)*

Country	Frequency of assessment	Funding tool	Indicators used	Budget affected
United Kingdom	RAE conducted every 5-7 years; annual activity metrics	Non-capital research block funding	The principle indicator is the peer review conducted under the Research Assessment Exercise (RAE). Other indicators include: 1. Charity income 2. Volume of business research (income) 3. Volume of postgraduate research supervision	The RAE plus metrics drive an allocation of funding for each institution that currently accounts for around one-third of non-capital public spend in science and research allocated through the dual support system (block funding plus project funding)

1. Student load refers to the equivalent full-time student load (EFTSL), which is a measure of the study load, for a year, of a student undertaking a course of study on a full-time basis.

Source: Country responses to OECD RIHR questionnaire on performance-based funding for public research in tertiary education institutions (February 2010).

Most countries appear to have fixed funding pools, with funds allocated according to relative performance. The majority make payments under their performance-based funding arrangements at the highest institutional level, even if the assessment is made at more disaggregated levels; payments are not made at the level of research centres within universities, individual researchers, research groups or networks. In Poland, however, assessment and decision on financing are made at the faculty level; funding goes to the tertiary institution, but the ministerial decisions on allocations to faculty cannot be altered by the university administration. In a few cases, caps and floors are used to limit the extent of budget gains or losses due to performance-based allocations – this issue is discussed below.

Making results public

Most countries have very open disclosure of processes and results of performance-based funding systems. In Australia, grant amounts for each programme and the methodologies for the calculation of funding are publicly available on the Department of Innovation, Industry, Science and Research (DIISR) website. Danish statistics on performance are released on the website of the Danish Agency for Science, Technology and Innovation. They include results at the level of institutions (universities), main research areas (arts, political and social science, natural and technical science, and medical science), and main research areas within each university. The results of New Zealand's PBRF are publicly released via a number of channels, such as reports on the quality evaluation exercises (including overall results, trends, processes and panel reports) and PBRF annual reports (with data on each institution). Results are published at the level of the institution, not the individual; however, individuals can know the quality category assigned to their portfolio of work, and institutions were given a confidential report following the 2006 quality evaluation that outlined the grades assigned to individual staff by the peer review panels. The provision of information is also subject to New Zealand's Official Information Act 1982 and the Privacy Act 1993.

Open processes are also found in Norway, where institutions receive an information booklet each year outlining the main changes in the budget, accompanied by the underlying documentation (performance on each indicator per institution). The results are also made public on the government's website and accessible on the website of the Database for Statistics on Higher Education (DBH). Finland's system is similar to Norway's – an information booklet on funding models and indicators is published and sent to universities and is made public on the Ministry of Education's website. Poland's performance assessment processes and results are available to the public on the website of the Ministry of Science and

Higher Education. Results of the United Kingdom's RAE are made public by unit of assessment within each higher education institution but are not disclosed for individual researchers in the exercise.

Austria noted that it publishes the results of the performance-based assessment to universities after the tri-annual university budgets have been established. Sweden noted that the assessment processes are generally known. Results of the performance-based allocations by German *Länder* are provided to HEIs and the regional parliament. In some states (Baden-Württemberg, Lower Saxony) they can also be requested by the public. In the Czech Republic, assessment processes and results are made public.

Internal allocation of performance-based funds

With respect to the allocation of performance-based funds within institutions, the results of the questionnaire indicate a scale of government control, ranging from decisions that are purely internal to the TEI, to government expectations of certain behaviour, to formal requirements on allocations:

- In Germany and Norway, institutions may internally allocate the grants as they wish, and are free to decide whether and to what extent the regional or national-level incentives (or any other incentives) are employed in their internal budget distributions. The situation is similar in New Zealand, where institutions are free to determine the allocation of the PBRF funds. In Australia, higher education providers are expected to use their own allocation mechanisms to distribute their research block grants internally. In Austria, some institutions also use indicators (on different levels) to allocate funds internally. In Denmark, universities are free to use their revenues according to their own institutional priorities and there are no government stipulations on how performance-based funds are distributed internally. It is assumed, however, that the government incentive structure influences internal incentive structures at the universities.

- In Finland, performance-based funds are allocated internally at the discretion of the university according to its strategic choices, although with reference to target outcomes agreed with the Ministry of Education. In Poland, institutions have freedom to allocate funds internally, within the framework of their annual applications for statutory funding, which set out planned research directions and which are implicitly an agreement between research units and the ministry on spending directions. In Sweden, institutions are expected to change their internal allocation of funding to create incentives to seek external funding and to publish good quality papers, although

this expectation is not explicitly set out in any regulations. In the United Kingdom, HEIs are autonomous in deciding the uses of funding, but they are expected to take strategic decisions on research funding based on their strengths and with a strong focus on research excellence.

- In Belgium, there is some government direction for spending performance-based funds. The BOF funding is allocated internally by the university board, on advice from the research board and according to selection criteria and allocation rules set out in the Ministerial Order on BOF. However, the IOF is allocated internally by the association board on the advice of the industrial research board. In the Czech Republic, the internal allocation of institutional performance-based funds must follow the general rules set out in the legislation (Act No. 130/2002 Coll. on the support of research and development) and the European Union document on state aid for research, development and innovation (2006/C 323/01). The targeted funding for specific university research is administrated by TEIs, following the general rule that the major part of the funds must go to student research.

Governance and the role of institutions in designing and administering systems

In most countries, TEIs have played some role in designing the performance-based funding system and a number continue to be involved in its administration via data collection and ongoing consultation. Levels of TEI involvement range from high to low, generally as follows:

- In Denmark, the new system of allocating the basic grant according to performance is based on a proposal from the association of Danish universities, which was developed through negotiations between universities and the ministry. In addition, representatives from the universities are involved in the Steering Committee, around 300 researchers are represented in the expert groups responsible for selecting the data (journals and publishers) to be included in the bibliometric indicator, and the association "Universities Denmark" is responsible for appointing Danish researchers to the expert groups. The Danish Agency for Science, Technology and Innovation is responsible for oversight of the system and its administration and the Danish University and Property Agency is responsible for the allocation and distribution of funds to the institutions.

- In Belgium, the BOF allocation key introduced in 2003 was proposed by the Flemish interuniversity council (VLIR) which also has a role in data collection and processing for indicators on scientific personnel. The Centre for R&D Monitoring (ECOOM), an interuniversity consortium in which all Flemish universities participate, collects and processes data on publications, citations, patents and spin-offs. The Flemish ministers competent for education and innovation are responsible for the oversight of the performance-based system and the distribution of funds.

- In Poland, the system was designed by scholars elected to the State Committee for Scientific Research (until 2003) or nominated to the Council of Science. Other bodies, such as rectors' conferences, have an impact on the form of the evaluation system in the context of "social consultations". The Ministry of Science and Higher Education is responsible for oversight of the system, funds allocation and distribution.

- In Norway, representatives from the higher education sector, the ministry and the Norwegian Association of Higher Education Institutions (UHR) collaborated on the design and implementation of the funding system. UHR also played an important role in the design of the publication indicator in 2004 and has since had responsibility for the maintenance and further development of the indicator and its database. This responsibility is carried out by a National Publishing Board with representatives at the level of deans from all types of institutions and major research areas. Tertiary institutions themselves do not have a formal role in the administration of the system, except to report their results to the Database for Statistics on Higher Education (DBH). The Ministry of Education and Research is responsible for calculations, funding allocation decisions and distribution of funds.

- In Austria, the public universities took part in the final phase of development of the indicator-based budget and are partly responsible for the collection of the base data for the indicators. The indicator-based budget is administered by the Austrian Ministry of Science and Research.

- In Finland, the implementation of performance-based funding systems was jointly developed by the Ministry of Education and the universities. The Ministry of Education is responsible for granting the formula-based core funding to universities for the execution of their statutory public duties, according to the extent, quality and impact of the activities and to education and science policy objectives.

- In the Czech Republic, both performance-based funding systems were drafted under the partnership and oversight of all stakeholders. Responsibility for ongoing oversight of the system, funding allocation decisions and the distribution of funds lies with the Council for Research, Development and Innovation, the Ministry of Education, Youth and Sport, and some other ministries.

- In the United Kingdom, HEIs are consulted in each round of assessment and when the system is changing. They also contribute experts to the assessment panels. The performance-based system is managed on behalf of the Department of Business Innovation and Skills, the Welsh Assembly, the Scottish Government and the Northern Ireland Assembly by each region's respective funding body. The level of funding is the responsibility of the respective administrations and the allocation of funding is the responsibility of the respective funding council or equivalent body.

- In the German states, HEIs are involved to varying degrees in the conception, administration and implementation of the systems, while the regional science ministries are responsible for the administration of the system. For example, in Baden-Württemberg, the system was developed and modified in co-operation between the Ministry of Science, Research and Arts and the universities and other HEIs. In Bavaria, the university associations co-operated with the Ministry of Science, Research and Arts on development and adaptation of the system, while in Lower Saxony the state conference of HEIs played a similar role. In Berlin, the state senate and TEIs developed the indicators, with some indicators based on the universities' suggestions. Brandenburg has a working group formed of members of TEIs and the Ministry of Science, Research and Culture, which monitors and develops the allocation system.

- In New Zealand, tertiary education organisations are responsible for providing complete and accurate data for the calculation of performance-based indicators. They have also been involved in the ongoing review of the PBRF system through consultation processes. The Tertiary Education Commission (TEC) is responsible for payment of PBRF funds and shares responsibility with the Ministry of Education for validating and verifying data. Monitoring and reporting within the system is based on high trust, high accountability and low compliance costs.

- In Australia, research block grant guidelines, conditions of grants and funding amounts are determined by the government Minister for Innovation, Industry, Science and Research (or delegate), but any

programme or policy change is usually subject to consultation with the university sector. The government consults with the higher education sector in determining how the allocation of other research block grants may be linked to ERA results.

- In contrast, in Sweden, tertiary institutions have no role in the design, administration or implementation of the performance-based funding system. The Ministry of Education is responsible for system oversight, fund allocation decisions and distribution of funds.

Effects of using performance-based funding

Most countries have not yet undertaken formal evaluations of their performance-based funding schemes, in part because a number of systems are still too new to identify specific impacts. For example, Austria noted that its indicator budget has been used only twice so far and that impacts and the effects on the performance of Austrian public universities are still being evaluated. Sweden is in a similar position, with its system only in place since 2009. However, the Swedish government will appoint an Inquiry Chair to review the current system to see if peer review and co-operation with the surrounding community should be components of the system. Any changes would likely appear in the next Research Bill, due in 2012. Belgium's IOF policy was introduced in 2005, but with a two-year time-lag on the formula's parameters, it is too early to see effects. With changes to systems in 2008/09, the Czech Republic plans to monitor results continuously from 2010 and evaluate impacts at a later date. Denmark plans to undertake small evaluations each year from 2010, with a larger evaluation due in 2013. Poland has not yet formally evaluated its instrument for institutional funding, but noted that scholars send their opinions to the ministry and assess the system in the press (notably in the *Forum Akademickie* journal). Furthermore, in late 2009, the Department of Strategy undertook a consultation concerning the value of the institutional funding system, and a report drawing on the 135 responses is being prepared.

However, some countries have undertaken more formal evaluations of their performance-based funding systems. In Australia, the 2004 evaluation of the 1999 policy framework, *Knowledge and Innovation: A Policy Statement on Research and Research Training*, touched on the performance-based IGS and RIBG policies. New Zealand established a three-phase evaluation strategy for its PBRF, with phase one covering implementation, phase two giving a sense of emerging trends (Adams, 2008), and phase three (to be undertaken after the 2012 evaluation round) providing a longer-term assessment of outcomes. The Ministry of Education has also published two studies on the influence of the PBRF on research quality. Norway's Ministry of Education and Research evaluated the Norwegian funding model in 2009;

the report was presented in the Annual Budget Report to the Norwegian Parliament. The evaluation aimed to establish a knowledge base on the effects of the funding system and to identify adjustments that would contribute to achieving the sector's goals. In the United Kingdom, several analyses of the impacts and consequences of the performance-based system have been undertaken in recent years. In Germany, Brandenburg's allocation model has been reviewed by external evaluators. In Finland, the university funding model was evaluated as part of the international evaluation of the Finnish national innovation system.

Evaluations and anecdotal evidence have pointed to a number of positive impacts arising from the use of performance-based funding, particularly in regard to research outputs but also research management. For instance, the Australian evaluation suggested that subjecting research block funding to performance formulas improved universities' strategic focus on research and their research performance. In Belgium, ECOOM, the body responsible for calculating the bibliometric indicators for the BOF scheme, found a stimulating effect on universities stemming from the bibliometric criteria for funding. The Czech Republic suggested that there were more research results, an improvement in publishing activity by students and young researchers, and integration of more research activities in tertiary education. It also noted more targeted investment in research equipment installed in TEIs, and an overall improvement in quality in tertiary education. Finland pointed to an increase in the number of new PhD graduates (doubling from 1993 to 2008) and growth in the number of researchers and international scientific publications.

In New Zealand, a comparison of PBRF indicators from the 2003 and 2006 quality evaluations suggested there had been an improvement in the average quality of research at participating institutions, and that the percentage of New Zealand researchers undertaking world-class research had also risen. In the formal evaluation of the PBRF in 2008, Adams (2008) concluded that there had been positive effects on the behaviour of individual staff and on the processes and mechanisms used to support research within institutions. He also considered that the introduction of the PBRF had the beneficial effect of reducing the emphasis on teaching and student numbers that had characterised the New Zealand tertiary system in the 1990s and that had placed research as an adjunct rather than core activity. Recent studies by the Ministry of Education have found an increase in the impact of research (as measured by citations) since the introduction of the PBRF. It is suggested that poor performance in the 2003 quality evaluation was a key factor in improved research quality in the 2006 round.

Norway's evaluation suggested that research incentives have made a significant contribution to the increase in the volume of Norwegian scientific publications, even though the performance-based publication indicator has a reasonably modest impact on funding (less than 2% of total TEI funding). The Ministry of Education and Research suggested that this may be partly explained by intra-institutional use of the publications indicator for the purposes of management, which may lift the visibility and impact of this indicator. (At this time, there is no available analysis on the impact of internal funding distribution models at national level.) The number of PhD degrees has also increased in Norway, although this is partly due to an increase in funded positions.

The United Kingdom noted that the introduction of the current performance-based system coincided with an inflexion point in the performance trend of the research base, with its share of citations growing faster than its share of world publications (Evidence Ltd, 2005). There was also an improvement in the performance and the concentration of resources in more research-intensive institutions. The introduction of overt science management systems within HEIs was believed to be linked to greater institutional autonomy in the use of funds and the drive towards excellence. One positive but unintended effect was the engagement of "higher quality academic staff" due to recruitment strategies targeting excellence. The survey response also noted qualitative evidence from vice-chancellors of HEIs that a highly selective system of allocating public funds attracts private research funding.

In Germany, the types of positive impact noted by HEIs and responsible ministries were slightly different, in line with the objectives of the schemes. The funding models were viewed as having positive effects on governance, particularly by stimulating debates on performance components, strategic steering and prioritisation, and institutional leadership within administrations, HEIs, faculties and departments. The transparency and rationality of grant allocations improved and budget negotiations were sped up through use of performance-based systems. States also identified increases in third-party funding and doctorate degrees awarded, and noted that HEIs had become more competitive.

However, in identifying the positive impacts of these systems, several countries noted the difficulty of isolating the effects of performance-based funding from those of the general funding environment. Belgium noted that it is difficult to attribute impacts explicitly to specific funds, as they form part of the broader university funding system. New Zealand suggested it would be an oversimplification to attribute improvements in quality, and increases in degree completions and external income, entirely to the performance-based system, and Germany expressed a similar view with respect to observed

increases in third-party funding and graduations. Norway noted that funding from the Research Council and the EU to the TEI sector has increased but that this cannot be unequivocally attributed to the incentive structure. The United Kingdom also noted the difficulty of concluding that changes in performance are due to a single factor.

The RIHR survey responses received from countries also highlighted some potentially negative and/or unintended effects of their performance-based funding systems, with one broad issue being a narrowing of research focus. In New Zealand, for example, private R&D-intensive firms have asserted that the PBRF acts as a disincentive for tertiary institutions to partner with, and transfer knowledge to, the private sector. They suggest that the system's focus on academic excellence deters tertiary-sector staff from working on more commercially focused, applied areas of research, and that the drive towards publication is a barrier to conducting commercial in-confidence research. The 2008 evaluation by Adams also raised questions about the way in which research in institutes of technology and polytechnics, which tends to be more practice-based, is dealt with under the PBRF and whether alternative processes are required to support their complementary (but different) missions. More generally, the system may encourage researchers to focus on a narrower range of outputs (*e.g.* preferring journal articles and publication in a narrower range of journals). The United Kingdom suggested there may be higher risk aversion in the choice of research areas (*e.g.* if the potential for citation is limited) and lower levels of multidisciplinarity as a result of its research assessment exercise. Germany's HEIs also expressed concerns that a dominance of quantitative indicators might lead to incentives that are detrimental to quality.

In New Zealand, another unintended effect of the PBRF has been increased incentive for tertiary institutions to compete with other research organisations, including New Zealand's eight Crown Research Institutes, for publicly funded research grants from the research, science and technology budget allocation (as these are counted as external research income). There may also now be an undue focus on staff with established track records, at the expense of creating a sustainable profile of staff age and experience across a department, because the quality assessment is conducted at the level of the individual. In the United Kingdom, there is a broader question, as yet unanswered, as to whether further concentration of resources is a desirable outcome and will support a world-class national research system (Adams and Gurney, 2010). The new weightings that drive funding from 2010-11 will concentrate resources in favour of the five top-performing institutions, contrary to the wishes of those who argue for more diversity. For their part, Adams and Gurney argue that the present level of concentration has supported

research that gains global recognition and disturbing this arrangement could be detrimental.

The German response to the RIHR questionnaire highlighted the importance of implementing indicator-based allocation mechanisms consistently. It noted that where systems had been supplemented by additional, non-transparent and discretionary measures (particularly in order to recognise qualitative efforts, *e.g.* on regional planning), the allocation system became less transparent and less predictable, which contradicted the aims of the schemes.

Indicators and incentives

The effects of performance-based systems rely to a great extent on the indicators used and the incentives they generate. Over time, most countries have "fine-tuned" their indicators so as to better target the desired goals. In Australia, for example, so as to reward universities that diversify their sources of income beyond Australian competitive grants, the JRE (replacing the IGS) will no longer include this variable in the calculations. In a slightly larger change, Australia also removed capping of payments under the IGS and replaced this with a safety net. Previously, the formula for funding the IGS had capped payments so that institutions could not gain more than a 5% increase in IGS funding above their previous year's allocation. Increases larger than 5% were redistributed to institutions that had suffered declines in funding. The 2004 evaluation criticised this, noting that capping of funding disadvantaged successful institutions and worked against the intent of policies. The new safety net mechanism means that now no grant can fall below 95% of the previous year's grant (indexed to current prices). In Germany, most of the models implemented by the *Länder* limit both the maximum possible budget loss and gain of a HEI by using a tolerance band linked to the total volume of the previous year's budget. In Berlin this band is 5%, while in other states it is between 1% and 2%. At the same time, the share of the budget determined by indicators has risen in several *Länder*. The 2008 evaluation of New Zealand's system touched on the issue of caps and funding impacts; Adams (2008) suggested that the overall funding envelope for research may need to be increased, if improvements in quality are to be rewarded (via salaries or other resources for research) and thereby drive the outcomes targeted by the policy.

Some countries have added indicators; for example, in Belgium, a mobility and diversity parameter was added to the indicator mix in 2004, with the weight gradually rising to 4% of the total, in order to encourage universities to hire women and externally trained researchers. In Berlin, the set of research-related indicators was expanded to reward participation in clusters of excellence or research centres. In Australia, a new performance-

based funding tool, the SRE, aims not only to assist with research costs but also to encourage the use of best practice financial management, performance and reporting frameworks. Countries also made improvements to their underlying data sources. For instance, in 2008 Belgium began to supplement Web of Science bibliographic information with other international bibliographic files. It also created a Flemish academic bibliographic file for social sciences and the humanities, which will be put into operation in 2011. In some cases, there has been an evolution from input measures to more output-based measures. For instance, Finland noted that the proportion of criteria directly or indirectly describing the quality of operations has increased to about one-third of the funding. In other changes, Norway unified in 2006 the indicators to cover all institutions, rather than differentiating between universities, specialised university institutions and university colleges.

Despite ongoing changes to indicators over time, there have been some notable criticisms of performance-based systems, particularly concerning the structure of incentives. In Norway, for example, some feedback has suggested that financial incentives to institutions to seek funds from the Research Council of Norway and the EU Framework Programme may be unnecessary. The low success rate in gaining such external funding means there are already substantial kudos for successful institutions so that encouraging applications may not be useful. Nevertheless, the Ministry of Education and Research considered that incentives for institutions to seek this type of external funding are important for motivating institutions and enhancing quality. Norway also stated that it was not evident that performance-based funding has prevented fragmentation of research funding. In Belgium, analysis of unwanted effects stemming from the BOF allocation key has raised issues about the appropriate influence of publications with lower impact factors, the lack of an international focus, and the domain-independence of the key (discussions on this analysis are ongoing).

Feedback received in Poland on its performance-based system for institutional funding highlighted a number of concerns about the operation of the system's indicators. Many research-related activities were considered to be under-rewarded or not rewarded appropriately (see Box 3.2). For instance, comments suggested that the system encouraged competition instead of co-operation, since participation in domestic or international research networks or consortia is not acknowledged.

There were also concerns about how well the Polish system dealt with field specificity. It was suggested that the evaluation regime should differ across units (*e.g.* condensed matter physics and particle physics should have slightly different regimes) and across disciplines (*e.g.* criteria should be different for disciplines closer to technological application). It was also suggested that there was not enough consideration of subject-based differences in the number

of publications, the rate and longevity of citations, and the importance of the peer-reviewed journals included in the Web of Science or Scopus list. In general, countries appeared to have a mix of levels of field specificity; for instance, the Danish bibliometric research indicator is applied equally to all disciplines and a key issue in its design was that it could fit all main areas of research equally. In Sweden, however, there is a weighting system to compensate the humanities for their lesser ability to attract external funding and for differences in publishing traditions compared, for example, to medicine and technology. The systems used by some German *Länder* (including Baden-Württemberg, Bavaria, Hesse, Lower Saxony and Rhineland-Palatinate) also apply subject- or discipline-specific weighting.

Box 3.2. Rewarding research activities: Polish feedback

Feedback on the Polish system for institutional funding has highlighted a number of areas in which the rewards for research activities may be under-rewarded or inappropriately rewarded given their level of merit. Examples of under-rewarded areas included:

- Small scale experimental production.
- Spin-offs and spin-outs.
- Success in grant and contract acquisition (including small contracts, since in many technological areas, co-operation with small and medium-sized enterprises (SMEs) consists of many small agreements).
- Science education for schools and the general public, organisation of public debates, and other scientific communication activities (*e.g.* maintenance of portals).
- Education of new researchers and support for mobility.
- Science and technology activities, such as measurement, monitoring, stocktaking and maintenance of data.
- Preparation of governmental regulations, instructions and standards.
- Translation of important handbooks or monographs of foreign authors; and
- Maintenance of accredited laboratories.

Examples of potentially inappropriate rewards included the favouring of English-language publications, rewarding publication quantity, not valuing certain accomplishments (such as participation in governing boards of international research organisations) according to merit, and greater rewards for some activities than for others (*e.g.* awarding 700 points for employing the laureate of the "Ideas" contest of the European Research Council, compared to three points for a chapter in a scientific monograph or academic handbook in languages other than English).

Source: Polish response to the OECD RIHR Questionnaire on Performance-Based Funding for Public Research in Tertiary Education Institutions (February 2010).

Future changes and developments

Several countries highlighted intended or desired changes to the system of indicators used in their performance-based funding systems, with some countries seeking to simplify their systems while others intended to add elements. In Austria, there have been calls to sharpen the indicators and simplify the mathematical model behind the system in order to make it easier to predict the impact on institutions. Belgium also noted scope to improve the transparency of its allocation key for the BOF, as the formula has become complicated. In Denmark, citations may be used to a greater extent in the future. Currently, journal impact factors are used voluntarily as a tool to rank journals. In Finland, the indicators will be further developed to describe education, research, research education and artistic activities. This will require consideration of the appropriate classification of publication data, promotion of the four-step researcher career[6] and research co-operation. Publication-based funding is likely to be given more weight as the classification of publication activity advances. In Norway, the 2009 evaluation foresaw no need for extensive changes to the funding system, but some minor adjustments will be made, with effect from 2012. These include incorporating regional research funds into the allocation formula and reducing the weighting of EU funding in the formula. In Poland, changes were planned in the project of the law on science financing that was recently accepted by the relevant commission of the Policy Parliament. It is expected that the system will move towards more differentiation of criteria (while maintaining common core rules) and more visits *in situ*.

At the level of whole systems, changes have been more limited, although redesign is in the pipeline in several countries. In the Czech Republic, the model for funding specific university research (student research) was made more targeted in 2008, but no further changes are envisaged. In New Zealand, two new quality categories were introduced for the 2006 evaluation round, so as to better recognise new and emerging researchers. Changes ahead of the 2012 evaluation are likely to involve the treatment of commercial research, and may also address concerns about the provision of individual scores to institutions (considered as potentially undermining staff development processes in some institutions). Of note, the quality scores yielded by the peer review quality evaluation process in New Zealand are considered to have been fairly low (an average score of 2.96 out of 10 for the 31 tertiary institutions participating in the 2006 evaluation round; see TEC, 2006). This raises the question of whether it is the actual level of quality or rather the design of the scoring process or the breadth of tertiary institutions included in the system that is driving results, and whether changes to the system are needed. In the United Kingdom, the grading system was changed in 2008 to provide each unit of assessment with a "research profile" that recorded the

percentages of research submitted that fell into each of four grades, ranging from world-leading to nationally recognised. This had the effect of better indicating quality and has led to broader distribution of funding. A consultation in 2006 about moving towards a system largely based on citations was rejected by the academic community. In future, however, there will be more recognition of the contribution of science and research to society, by including an explicit element on the impact of excellent research within the new Research Excellence Framework (REF) that is currently being developed. Poland also plans to put more stress on the measurement of social and economic impacts of research.

In other changes, the Australian evaluation found a need to strengthen the drivers for excellence in research and for linkages and collaboration between universities and the national innovation system. The SRE scheme, which will receive AUD 510 million from 2009/10 to 2012/13, primarily aims to help universities address shortfalls in funding for indirect costs associated with competitive grants, but is allocated according to performance (in particular, using the results of the new ERA system). The introduction of the performance-driven JRE scheme in 2010 is aimed at encouraging greater collaboration between universities and the business and non-government research sectors by rewarding higher education providers that diversify their sources of income. Finland suggested that the level of internationalisation could be improved via changes to the performance system.

The German response to the RIHR questionnaire noted that care must be taken when changing funding models, as this can create confusion and gaps in ongoing governance arrangements. Higher education commissions have been set up in many German *Länder* to help develop allocation models, and have made recommendations for changes and improvements. Most of these have maintained continuity with state governance strategies, but in Lower Saxony and Berlin it was considered necessary to replace the funding systems. The German response noted that large changes can potentially cause "fissures in the steering logic", and as universities and other HEIs often link their internal funding systems to state-level systems, frequent changes are not recommended.

Interactions with the performance-based funding system

Performance-based funding systems for public research in TEIs do not operate in isolation. The OECD RIHR questionnaire sought information on the interactions of these systems with funding systems for teaching, other evaluation systems and activities, and other funding mechanisms. The information provided also highlighted the role of these funding mechanisms within the broader governance structures for TEIs.

Funding of teaching

A number of countries have incorporated performance-based aspects into their funding of teaching at TEIs but these were not regarded as problematic for the operation of the research funding system. For instance, Austria's performance-based system encompasses both research and teaching components, and Finland's core funding for universities incorporates a performance-based allocation for education, which takes into account the extent of activities and their quality and effectiveness. The Czech system for funding teaching comprises several components, one of which is performance-based and uses indicators such as number of students, financial assessment of accredited programmes and results achieved in activities undertaken. Germany's institutional funding system makes no distinction between research and teaching, and the performance-based allocation systems used by the *Länder* thus encompass both aspects, with no identified frictions. In Norway, part of the total funding for tertiary institutions (around 24%) is distributed on the basis of performance in education, and in particular, results in the areas of education quality and internationalisation. The grant is distributed according to the performance two years earlier in the number of study points obtained by students and the number of incoming and outgoing exchange students. The education incentives have fixed rates for each indicator, meaning that improved results for one institution will lead to increased allocations regardless of the performance in the rest of the sector (*i.e.* the funding pool is not fixed). There is no particular interaction between the two funding mechanisms of teaching and research and institutions are free to use the performance-based funding received for education as they wish (*e.g.* it may be used to fund research). The United Kingdom has a parallel assessment system for teaching, but it has no interaction with the research system.

Several countries are in the process of incorporating more performance-based assessment into teaching funding decisions. For example, Australia currently does not have a performance-based funding model for teaching in universities (a 2006 Learning and Teaching Performance fund aimed at rewarding universities which best demonstrated excellence in teaching and learning was discontinued). However, from 2012 the Australian government will introduce new performance funding (AUD 206 million over four years) to reward universities that meet agreed targets in key areas including improving the quality of learning and teaching and outcomes for students from low socioeconomic backgrounds. Interaction is not expected between this at-risk funding for universities and the performance-based block grants for research and research training. In New Zealand, a performance-based funding system to support educational performance in the tertiary sector is currently under development. It will make some funds subject to educational achievement

targets (such as qualification and course completions and transitions to further education), but is not expected to specifically interact with the PBRF. In Sweden, the government presented a bill in early 2010 that proposed rewarding universities receiving high appraisals in quality evaluations based on three criteria: students' degree projects, self-evaluations of HEIs in combination with site visits, and questionnaires sent to alumni.

Other countries do not use performance-based funding for teaching. Belgium's funding of teaching in universities is based on the number of credits taken by students, in addition to other factors. Denmark does not have a performance-based model for teaching. In Poland, the funding of teaching is predominantly metrics-based and is regulated by a different law.

Other evaluation exercises

A number of countries identified other research evaluation exercises and, from the information provided, it does not appear that these evaluation exercises clash with the performance-based funding system. However, countries should check periodically that the exercises continue to be complementary and that their results and recommendations are not imposing contradictory demands on TEIs. Examples of other evaluation exercises include:

- In Austria, the major annual statistics reporting system for public universities (*Wissensbilanz*) includes a number of research indicators, such as personnel numbers in research programmes, publications, patents, external presentations, scholarships and clinical studies.
- Czech ministries conduct an evaluation of all measures focused on research funding (programmes, projects, grants, final results), according to the Act No. 130/2002 Coll.
- Flemish universities organise an external evaluation of the quality of their research management in general, and of the operations of their research boards in particular, every eight years. The results of these evaluations are made publicly available. The government also regularly evaluates research councils, strategic research centres and other R&D partners that receive public funding.
- Finland noted that the Finnish Higher Education Evaluation Council (attached to the Ministry of Education) and the Academy of Finland undertake regular evaluations and external audits of quality assurance systems in universities, to enhance the quality and impact of their activities. The Academy of Finland and the Finnish Funding Agency for Technology (Tekes) are currently undertaking a project to construct a model for systematic and permanent evaluation of the impact and quality of research that they fund.

- New Zealand's Ministry of Research, Science and Technology (MoRST), the Foundation for Research, Science and Technology (FRST), the Ministry of Education, and the TEC all conduct evaluations of publicly funded research outcomes. The focus is generally on the value of specific policy interventions rather than the value of the research itself.

- In Norway, the Norwegian Research Council conducts national-level subject-specific evaluations of research to provide a critical review of the research system in an international perspective and recommendations to increase the quality and efficiency of research. On average, two subjects are evaluated each year. Also, in the context of quality assurance exercises, the Norwegian Agency for Quality Assurance in Education (NOKUT) may consider the research capacity of the institutions and subjects evaluated and may include research-relevant proposals in national development plans for subjects.

- There are a number of evaluation exercises in the United Kingdom, including periodic reviews by research councils of the quality and effectiveness of their support in specific areas or disciplines and monitoring reports and surveys from the Department of Business Innovation and Skills on the use of resources and the outcomes and outputs of research council investments.

Australia, Denmark and Poland noted no other national systems of regular evaluation of research conducted by ministries, research councils or other bodies. Several countries noted systems of *ex ante* evaluation of research proposals by researchers and research groups.

Other funding

The wide range of research funding channels for TEIs may raise the possibility of tensions with performance-based funding systems, particularly if co-ordination processes are not in place. The Australian survey response pointed to the potential for friction and co-ordination problems when research funding comes in several different streams (block grants, project funds) provided by different government departments and, in its case, different levels of government (commonwealth, state and territory). For example, other government departments need to work with the DIISR to ensure that undue pressure is not put on the performance-based RIBG scheme when introducing new competitive grant programmes for research (since the RIBG aims to meet project-related infrastructure costs associated with Australian competitive grants). For its part, Australia has interdepartmental co-ordination processes in place for major policy development that help to

minimise friction and ensure complementarity of policies. There are also formal bodies, such as the Prime Minister's Science, Engineering and Innovation Council, which aid with governance and co-ordination at the national innovation system level. However, Poland noted that while there are no visible frictions between different funding instruments, there is unfortunately also no co-ordination between project and institutional funding provided by the Ministry of Science and Higher Education and between the funding of the Ministry and the Foundation for Polish Science. In the United Kingdom, there is no explicitly imposed co-ordination mechanism, but many elements of the funding system are interlinked and are developed jointly by the responsible funding bodies. The Czech Republic noted that synergies between different types of funding are monitored, and double- or cross-financial usage is legally forbidden.

In Germany, research has suggested that the "discretionary-incrementalist" components of funding alongside performance-based components have the potential to undermine or neutralise the impact of indicator-driven funding (Orr and Jaeger, 2009). Discretionary-incrementalist funds can be characterised as "single-case", non-transparent decisions that determine funding levels (*e.g.* a decision to allocate funds equal to the previous year plus a discretionary increase), and this type of funding remains significant in Germany for grant allocations. As such, traditional and new steering mechanisms must be better linked, so that "single-case" decisions do not undermine the budget-related effects of the performance-based schemes.

Some countries identified frictions between the performance-based system and external sources of funding. Norway's tertiary sector has suggested that indicators based on funding from the EU and the Norwegian Research Council bring challenges, namely that TEIs find it difficult to pursue research strategies not directly in line with research areas covered by the EU and the Council, and that the co-funding required with such project funding means less liberty for TEIs in allocating their block grants. Slovenia also mentioned the funding mechanisms provided by the EU Structural Funds, noting that while the funding brings significant developments (*e.g.* it gives rise to new forms of co-operative research and strengthens intermediary institutions such as technology centres), the complexity of the administrative procedures for this financing weakens the efficiency of these interactions and needs simplification.

The role of performance-based funding in the overall governance mechanisms of TEIs

As a tool for governing and steering research activities in TEIs, performance-based funding systems sit alongside (and, ideally, complement) other policy mechanisms within each country's higher education and research environment. For instance, in Australia, the Powering Ideas Innovation Agenda announced in the 2009-10 budget will see the introduction of mission-based compacts with universities, which will define each university's particular mission and describe how it will be fulfilled. The mission-based funding compacts, to take effect from 2011, are aimed at both capacity building and accountability, and will allow universities to determine their own research and collaboration agendas in line with national priorities. They are also expected to promote collaboration by encouraging universities to organise themselves into research hubs and spokes and to pursue opportunities to undertake more industry-driven research (Commonwealth of Australia, 2009). The compacts will facilitate the distribution of performance-based funds by defining targets for improvement and reform that will trigger reward payments.

Norway noted that the performance-based funding system is one of several tools used by the Ministry of Education and Research in its management of the higher education sector. In general, the *Act relating to Universities and University Colleges* gives institutions academic freedom and authority to determine their internal structure, while the ministry sets general goals and management parameters for the institutions as well as performance requirements in chosen areas. The system as a whole aims: to facilitate stable and long-term funding of universities and university colleges so that boards can set long-term strategies; to enhance the quality of education and research; to provide the impetus for efficient resource utilisation; and to allow institutions flexibility to adapt to the needs of society in accordance with their strategies. The funding system supports the goals for the sector as set out in the Act and specified in the annual national budget.

In Finland, the funding model is seen as enabling universities to undertake long-term development and encouraging them to develop distinct profiles, quality and productive and cost-effective activities. The funding criteria were designed to be as clear and transparent as possible so that future funding can be predicted to a reasonable degree. In the United Kingdom, the performance-based funding tool of the quality-related (QR) research block funding to institutions is part of the "Dual Support" mechanism of public-sector funding for research in HEIs. The second part of the mechanism is project-based funding via the research councils.

In some cases, however, other steering mechanisms may work against the goals of performance-based funding. In Germany, state initiatives to restructure the higher education system, by merging or closing institutions to reduce numbers of HEIs, were viewed as working against the autonomy of universities and other HEIs. Strengthening autonomy and institutional responsibility were core aims of performance-based allocations in the German *Länder*.

Key findings and policy issues raised by the questionnaire responses

The OECD RIHR questionnaire yielded information about a variety of national- and regional-level performance-based funding systems for public research in TEIs. Most current schemes have been introduced since 2000, although some have much longer histories. The rationales for these schemes generally include raising the quality of research, but many countries have additional goals, such as accountability and autonomy of TEIs, collaboration and technology transfer, and increasing critical mass (or, in contrast, for some countries, spreading funding more widely). Some countries do not have research-specific rationales for their schemes, as they also encompass teaching and other activities.

The funding systems generally operate on an annual basis, although some assessment processes (notably peer reviews) are conducted less frequently. Around two-thirds of schemes cover all TEIs, while one-third cover only public establishments, but all schemes cover all types of research and fields of activity. In the majority of cases, the systems affect a tranche of institutional funding, although some schemes operate via more disaggregated funding policies. Generally, funding is from a fixed pool, of which institutions receive a share based on their relative performance. There is open disclosure of processes and results in most countries, and institutions are relatively free to allocate funds received from performance-based systems as they wish.

The range of indicators used in the schemes reveals both similarities among countries and clear differences. They are similar in that third-party income, publications, student numbers and degree completions are commonly used, but different in that various combinations and weightings are employed. Some countries add many more indicators to the mix (*e.g.* patents, spin-offs), some do not use bibliometrics at all, some use peer review as a major or complementary component, and others are based purely on quantitative indicators. The weightings used for each item vary quite significantly.

The dimensions that differ most between countries, in addition to the indicators, are the budget impact of the schemes and the role of institutions in their development and administration. The questionnaire showed that different countries use different measurements and nomenclature in funding,

and it is difficult to compare how much institutional funding is performance-based. From the numbers, the range may be anything from 6% to 75% of annual block funding. It appears that a bigger allocation of performance-based funding is associated with the use of more indicators. The level of involvement of TEIs also ranges from being closely involved in scheme design and data collection to no role at all.

There have been few formal evaluations of performance-based funding schemes, but the evidence suggests positive effects on research outputs and research management. At the same time, it is difficult to attribute these effects directly to the performance-based schemes, as they are just one piece of the funding environment. Negative and unintended effects were also highlighted, with concerns about a narrowing of research focus as academic outputs, journal articles, certain fields and certain journals are favoured in more risk-averse environments.

Over time, indicators have been fine-tuned, with changes to weightings and design. However, there remain criticisms, including whether certain indicators are relevant, whether they provide appropriate rewards, and whether they are suitably field-specific. Future changes include both simplification and added complexity, and some countries intend wider system redesign, notably the United Kingdom. Interactions with teaching allocations and other evaluations were considered unproblematic, but the variety of funding streams for research was seen to require co-ordination in order to avoid frictions.

Some interesting policy issues that emerge from the questionnaire results include:

- How effective is it to have multiple goals for performance-based funding systems – can they all be achieved?
- What are the merits of seeking critical mass *versus* funding dispersion (both outcomes were sought in the studied systems)?
- Which indicators have the greatest impact on research incentives and which are peripheral to research decision making?
- What is the required amount of budget impact to provide incentives to TEIs? In some systems, the amount at stake is low, and some systems have caps or bands to limit the size of the budget impact. Gauging the effect of different levels of budget impact requires more accurate attribution of observed outputs and outcomes to the features of funding systems.

- To what extent is there an "ideal system", given that the questionnaire shows some countries moving towards more complex systems and others moving towards simplification?
- Should funding systems for teaching be more entwined with funding systems for research, given the dual purpose of TEIs?
- How should government co-ordination mechanisms be arranged so as to avoid frictions between different funding tools and channels?

Acronyms

BOF	Special Research Funds (Flemish Community of Belgium)
COBISS	The Slovenian national bibliographic system
DBH	Database for Statistics on Higher Education (Norway)
DFG	German Research Foundation
DIISR	Department of Innovation, Industry, Science and Research (Australia)
ECOOM	Centre for R&D Monitoring (Flemish Community of Belgium)
EFTSL	Equivalent full-time student load
EIT	European Institute of Innovation and Technology
EP	Evidence portfolio
ERA	Excellence in Research for Australia
ERI	External research income
ERIH	European Research Index for the Humanities
EU	European Union
FRST	Foundation for Research, Science and Technology (New Zealand)
GBAORD	Government budget appropriation or outlays for R&D (see OECD, 2002, pp. 137-150)
GDP	Gross domestic product
GUF	General university funds (see OECD, 2002, pp. 158-169)
HDR	Higher degree by research
HEI	Higher education institution
HEP	Higher education provider
IGS	Institutional Grants Scheme (Australia)

IOF	Industrial Research Funds (Flemish Community of Belgium)
ISBN	International Standard Book Number
ISSN	International Standard Serial Number
JCR	Journal citation report
JRE	Joint Research Engagement (Australia)
MoRST	Ministry of Research, Science and Technology (New Zealand)
NESTI	OECD Working Party of National Experts on Science and Technology Indicators
NOKUT	Norwegian Agency for Quality Assurance in Education
PBR	Performance-based reallocation (Norway)
PBRF	Performance-Based Research Fund (New Zealand)
QE	Quality evaluation (New Zealand)
QR	Quality-related (United Kingdom)
RAE	Research Assessment Exercise (United Kingdom)
R&D	Research and development
RDC	Research degree completions
REF	Research Excellence Framework (United Kingdom)
RIBG	Research infrastructure block grants (Australia)
RIHR	OECD Working Party on Research Institutions and Human Resources
SICRIS	The Slovenian current research information system
SRA	Slovenian Research Agency
SRE	Sustainable Research Excellence in Universities (Australia)
TEC	Tertiary Education Commission
TEI	Tertiary education institution
Tekes	Finnish Funding Agency for Technology
UHR	Norwegian Association of Higher Education Institutions
UK	United Kingdom
VLIR	Flemish inter-university council

Notes

1. The terms "institutional funding" and "project funding" are not specifically defined in the *Frascati Manual* (OECD, 2002), although the Manual notes that universities draw on R&D contracts and earmarked grants, as well as part of the general grant they receive from the ministry of education or corresponding local authority (known as "public general university funds" or GUF), in order to finance their R&D activities. These financial flows are in addition to universities' "own funds" which they may derive from endowments, shareholdings, property, student fees and so on.

2. A new Universities Act came into effect in 2010 in Finland, enlarging universities' financial and administrative autonomy by giving the former state universities the status of independent corporations under public law or foundations under the Foundations Act.

3. *Wānanga* are a type of New Zealand tertiary institution that provides education in a Māori cultural context. As defined in the Education Amendment Act 1990 s162(b)IV, *wānanga* are characterised by teaching and research that maintains, advances, disseminates and assists the application of knowledge regarding *ahuatanga Māori* (customs and material and spiritual objects and concepts) according to *tikanga Māori* (customs and protocols) (Ministry of Education, 2003).

4. See *www.oecd.org/document/3/0,3343,en_2649_34293_44904003_1_1_1_1,00.html*.

5. A 2004 survey of German universities found that 91% of universities in *Länder* with indicator-based allocation systems also used indicator systems internally, with 30% of these universities closely following the state-level model and 56% partially following it.

6. A working group under the Finnish Ministry of Education proposed a four-step research career model in universities that would promote transparency and predictability of research careers and would support movement of researchers between universities and other research actors (*e.g.* research institutes). The working group recommended that the Ministry of Education take into account the implementation of the four-stage research career system in universities when considering their funding and management by results. See *www.minedu.fi/OPM/Julkaisut/2008/Neliportainen_tutkijanura.html?lang=en*.

References

Adams, J. (2008), "Strategic Review of the Performance-Based Research Fund: The Assessment Process", report prepared for the Tertiary Education Commission by Evidence Ltd, June.

Adams, J. and K. Gurney (2010), *Funding Selectivity, Concentration and Excellence – How Good Is the UK's Research?*, report by the Higher Education Policy Institute (HEPI), United Kingdom.

Commonwealth of Australia (2009), *Powering Ideas: An Innovation Agenda for the 21st Century*, Canberra.

Evidence Ltd (2005), *Impact of Selective Funding of Research in England, and the Specific Outcomes of HEFCE Research Funding*, a desk-based review for HEFCE and the Department for Education and Skills, September, United Kingdom.

Jongbloed, B. (2010), "Funding Higher Education: A View across Europe", Report for the Modern Project: European Platform Higher Education Modernisation, European Centre for Strategic Management of Universities (ESMU).

Lepori, B., P. van den Besselaar, M. Dinges, B. Poti, E. Reale, S. Slipersæter, J. Thèves and B. van der Meulen (2007a), "Comparing the Evolution of National Research Policies: What Patterns of Change?", *Science and Public Policy*, Vol. 34(6), July, pp. 372-388.

Ministry of Education (2003), *Māori Tertiary Education Framework*, report by the Māori Tertiary Reference Group, November.

OECD (2002), *Frascati Manual: Proposed Standard Practice for Surveys on Research and Experimental Development*, OECD, Paris.

Orr, D. and M. Jaeger (2009), "Governance in German Higher Education: Competition versus Negotiation of Performance", in J. Huisman (ed.), *International Perspectives on the Governance of Higher Education: Alternative Frameworks for Coordination*, New York/London.

Steen, J. van (2010), "Better Use of Existing Data on Public Funding of R&D", paper presented at the 3rd ENID (European Network of Indicator Designers) conference on STI Indicators for Policymaking and Strategic Decisions, 3-5 March, Paris.

TEC (Tertiary Education Commission) (2006), *Performance-Based Research Fund: Evaluating Research Excellence: The 2006 Assessment*, Wellington.

Chapter 4

Impacts of performance-based research funding systems: A review of the concerns and the evidence

by

Linda Butler
Australian National University and University of Newcastle, Australia

This chapter examines the impacts of performance-based research funding systems. It discusses the difficulties of measuring impacts, particularly in distinguishing between intended and unintended consequences and in establishing whether outcomes are desirable or not. It also explains why presenting an evidence-based assessment is a challenge.

Introduction

Assessing the impact of performance-based research funding systems (PRFS) is a fraught exercise, which perhaps explains the paucity of broad authoritative texts on the subject. The literature is full of words like "likely", "potential", and "possible", but contains relatively few concrete examples that examine the impact of PRFS in detail, either through investigative data analysis or well-structured survey/qualitative investigations.

This deficit is acknowledged in *The Changing Governance of the Sciences* (Whitley and Gläser, 2008), which aims to:

> *"… set the stage for sociological research on RES [research evaluation systems] by outlining what is, and what should be, known about this major institutional innovation in the science policies of many countries. The contributions propose what is sociologically important about these RES and sketch the ways in which these phenomena could be investigated."* (Gläser, 2008, p. 246)

Policy analysts and researchers face a number of conundrums when attempting to identify responses to PRFS:

- How does one attribute causality? (Gläser *et al.*, 2002, p. 20)
- Is it the specific system that leads to particular responses, or simply the existence of an assessment regime, *any* assessment regime? Both peer review and formula-based systems can lead to increased output, institutional strategising, etc. (Gläser, 2008, p. 257)
- Are the changes in behaviour positive or negative? This can depend on the stakeholder group, the discipline, etc. For example, the debate on "transfer markets" in leading researchers just prior to each Research Assessment Exercise (RAE) cycle can be regarded either as game playing, or as driving the sector to provide salaries that are just rewards for achievements.
- What are the mediating effects of parallel reward systems? Requirements for tenure, appointments, promotions, targeted project funding, etc., all have their own requirements, which can be at odds with the signals sent by PRFS and can lead to varying responses in different countries, even for basically similar systems (Whitley, 2008, p. 24).

Yet research is seen as one of the major features determining an institution's reputation (European Commission, 2010, p. 56), and an understanding of the impact of any new, and existing, policy instrument is vital for its effective operation.

When the UK Department for Education and Skills (DfES) called for responses to a consultation document on a proposed reform of higher education research assessment and funding in 2006, it found that the question on "possible undesirable behavioural consequences" attracted the most comments, both from answers to the question directly related to the issue, and in more general statements. These implied that perverse behaviours were likely, whatever the system used (metrics or peer review) (DfES, 2006, para. 41).

This chapter consists of five sections. Following this general introduction, the intended consequences of PRFS are briefly surveyed, as this topic is covered in greater detail in Chapter 1 of this volume. The following section draws out existing evidence on the most common impacts generally regarded as unintended outcomes that are attributed to PRFS. Next, some related issues are touched upon, followed by a final brief conclusion.

A number of important distinctions are made throughout the course of this review. It covers both the intended, and the unintended, consequences of the introduction of PRFS across a range of countries. If the research sector responds to PRFS such that the government's underlying objectives are achieved, there can be little argument that this is positive. As Chapter 1 covers in detail the rationales behind the introduction of various systems, the main focus of this chapter is their unintended consequences. Actors in the sector respond in a variety of ways, and many responses are not anticipated by those designing the systems. It is these behavioural changes which are to be examined: What are they? And are they positive or negative?

The chapter also tries to distinguish between reality and perception. Many commentators on the "evils" of a particular system (and it is usually the opponents of a system who are most vocal) are merely postulating what they believe has, or could have, occurred, rather than presenting evidence on what has in fact occurred. McNay's study of the RAE demonstrated this clearly: while the number of staff who reported that they had moved away from interdisciplinary work was relatively small, almost half of those surveyed felt that the RAE hindered interdisciplinary work (McNay, 1998, p. 20). There was no doubt that many academics perceived the UK system to favour discipline-based research, but alternative policy drivers outside the RAE were clearly ensuring that interdisciplinary work continued to thrive.

A distinction is also made between comments that are evidenced-based and those that are purely anecdotal. This raises one of the most difficult aspects of this review. Given the importance of the issue, very little research has sought to identify and study responses to the introduction of PRFS. Gläser and colleagues have been particularly critical of the failure of researchers in science and technology studies (and the sociology of science) to study the effects of these systems (Gläser, 2008, p. 246; Gläser *et al.*, 2002, p. 3). There are relatively few studies to which one can refer for evidence:

- Research Information Network (RIN)/Joint Information Systems Committee survey of 800 UK academics (RIN, 2009).
- A study for the Higher Education Funding Council of England (HEFCE) – focus groups of senior academics and administrators, survey of heads of departments or research units, staff survey, interviews with limited range of users and funders of research (McNay, 1998).
- A second study for HEFCE focusing on the effect of the Research Assessment Exercise on interdisciplinary research, which surveyed 5 505 researchers (37% response rate) and 327 RAE panel members (62% response rate) (Evaluation Associates Ltd, 1999).
- The Roberts' review of the RAE which encompassed 420 consultation responses; 44 consultative meetings, six open public meetings; and nine workshops (Roberts, 2003).
- A Norwegian bibliometric analysis (Sivertsen, 2008).
- An Australian bibliometric analysis done in the Research Evaluation and Policy Project (REPP) at the Australian National University (Butler, 2003).
- A review of the impact of league tables based on a survey of 134 heads of institutions, with a 68% response rate (92) (CHERI & Hobsons Research, 2008).

It is clear that the bulk of evidence is based on the United Kingdom's RAE. This is not surprising given its much longer history, nor is it of great concern, even when, as here, the focus is on a much broader range of PRFS. As Gläser notes, vastly different systems can have quite similar impacts (Gläser, 2008, p. 256). Actors can only respond to the stimuli they receive in certain ways.

Both institutional and individual responses to PRFS will be examined here. Those responses are not necessarily the same, even within a single institution, as messages from university management can be filtered through several intermediaries before they reach individual researchers. Even when they are the same, researchers may well receive contradictory signals as they seek to obtain external funding for their research or face promotions and appointments committees.

Some literature on university rankings is also covered. While these have little direct impact on individual researchers, institutions respond strongly. It is, to a certain extent, a funding response: rankings equate with prestige, which in turn enables universities to attract more students and more interest from end users. Both translate into more funding.

A number of evaluation systems are meant to cover both universities and other types of research institutions, such as government research agencies (Australia), and all other research institutions (France) (European Commission, 2010, p. 96). However, this review shows that the overwhelming majority of the literature, and specifically studies with an evidence base, focus squarely on higher education institutions.

Intended consequences

Some common themes run through the rhetoric around the introduction of PRFS. All systems, of course, are seen as a means for selectively distributing research funds. But most also seek to use it to drive particular behaviours, most commonly an improvement in the quality of research undertaken; or to increase accountability on the expenditure of taxpayers' money (Frölich, 2008). In all cases, the goal of creating a funding allocation mechanism is achieved. However, the degree to which the funds are effectively distributed, and the extent to which secondary goals are achieved, is not always readily apparent.

The nuances of desired outcomes are apparent in the stated aims of PRFS across countries. The University Grants Committee of Hong Kong, China, lists three main outcomes it hopes to achieve (UGC, 2006, p. 3):

- role differentiation in the UGC sector,
- international competitiveness,
- "deep collaboration".

The UGC believes it has been effective in informing funding, symbolising public accountability and inducing improvement in research, but provides no evidence to support these assertions (UGC, 2006, p. 4). Indeed, the UGC acknowledges that the earliest research assessment exercises in 1993 and 1999

had "difficulty in distinguishing achievement at the top-end through a broad-brush exercise such as the RAE", and had taken out of the funding pool a small proportion of the block grant to reward the highest quality research. How this was done was not specified (UGC, 2006, p. 5).

The current Australian system is purely a funding allocation system, with no explicit subsidiary aims along the lines espoused in other countries. This may well arise from its early introduction and the acknowledgment of its deficiencies as a research assessment tool, which has seen alternative systems proposed. The objectives of the most recent initiative, Excellence in Research for Australia (ERA) are to:

- Establish an evaluation framework that gives government, industry, business and the wider community assurance of the excellence of research conducted in Australia's institutions.
- Provide a national stock-take of discipline-level areas of research strength and areas in which there is opportunity for development in Australia's higher education institutions.
- Identify excellence across the full spectrum of research performance.
- Identify emerging research areas and opportunities for further development.
- Allow for comparisons of Australia's research nationally and internationally for all discipline areas (ARC, 2009, p. 6).

As yet ERA has no direct funding implications, but institutions are responding on the assumption that this will change. In the interim the outcomes of the assessments will have significant implications for their ability to attract external funding and students, particularly full-fee-paying international students.

In some cases, existing measures were modified to stimulate the desired changes. Belgium's Flemish government included bibliometric data in its funding formula with a view to improved research performance (Debackere and Glänzel, 2004, p. 265). The formula will be further refined in 2011 by incorporating data from the social sciences and humanities along the lines of the Norwegian model (Sivertsen, 2010, p. 22).

Many countries are trying to achieve additional outcomes. For example in Sweden, policies have aimed at increasing the share of staff with doctorates. Their new model for distributing general university funds (GUF) includes an element that counts for 5% of the total – the number of staff with doctorates (Elzinga, 2009, p. 6).

Impact

The impact of a PRFS is multifaceted, and a number of actors influence final responses – institutions, managers and researchers themselves. In addition, many intervening factors can mitigate expected outcomes. Researchers may encounter contrary signals from the rewards system in their discipline, or other government agencies may run programmes that soften the impact of a relatively blunt PRFS.

The following teases out the evidence, or highlights the lack of it, in relation to the major consequences attributed to PRFS.

On funding

To date, the New Zealand government has undertaken the most detailed analysis, with published data, on the funding implications for individual universities of its decision to introduce its Performance Based Research Funding Scheme (PBRF) (Ministry of Education, 2008). In contrast, much of the analysis of the impact of the RAE on funding came not from government, but from the print media, most notably *The Times Higher Education Supplement* (*THE*). The funding implications of these two systems are discussed below, followed by a brief discussion of the compounding influence of international university rankings.

New Zealand's PBRF

The introduction of the PBRF had two main impacts on funding allocations: a shift in research funding away from institutes of technology and polytechnics (ITP) to universities, and a shift of funding to high-cost sciences at the expense of lower-cost fields (unspecified, but presumably social sciences and humanities). This followed a period during which increasing research funding flowed to the ITPs under the previous funding system. The ITPs' share of research funding was estimated to fall from 7.4% under the former scheme, to just 2.1% under the PBRF. Even though the funding pool was larger in the PBRF than the old system, ITPs collectively still received less than they would have under the earlier scheme. It was estimated they lost 65% of their research funding.

At the institutional level, Otago (26% increase in funding) and Lincoln (35% increase) were the big winners, while Auckland University of Technology, a relatively new university, received only half the funds that would have come its way under the old scheme.

In New Zealand's PBRF, fields are given differential weights to reflect the cost of undertaking different types of research. In the two evaluations conducted to date, the government has noted that this appears to have changed the fields that produce the most research income for institutions. Rather than low-cost fields with high student numbers, which previously drove the funding allocations, universities now derive more of their funding from high-cost fields (predominantly in the sciences) with relatively lower student numbers (New Zealand, Tertiary Education Commission, 2010, pp. 3-4).

The United Kingdom's RAE

Successive RAEs increasingly concentrated research funds in a few institutions, particularly the research-intensive Russell Group, until the 2008 exercise. Changing the assessment process from a single "star" rating for a department to a quality profile resulted in some major funding shifts (Corbyn, 2009). David Sweeney of the HEFCE, which is responsible for the oversight of the RAE, calculated a 3% drop in the concentration of research. For example, there were 25 institutions that received no mainstream quality-related (QR) funding in 2008-09 but which received funding in 2009-10. Many other institutions saw dramatic increases in funding at the expense of the research-intensive universities. For example, the University of Lincoln will get GBP 1.6 million in 2010 compared to the GBP 220 000 it received in the previous year. The Russell Group's share dropped from 65% to 60%, the first decline in its share of RAE funds.

However, in the most recent RAE, which introduced a new assessment methodology, there was considerable criticism of the weightings originally given to each star quality level, and the implied rewards for an improvement in performance from one level to the next. Adams and Gurney (2010) have undertaken a detailed analysis of the implications of RAE 2008 for funding selectivity. Under previous RAEs, institutions gained significant benefit from achieving the top ranking, compared to the second highest. This was not the case in 2008. The weights originally given to each level were 0 for 1*, 1 for 2*, 3 for 3*, and 7 for 4*. In effect, this means that a 3* was worth 3 times a 2*, but a 4* (a significantly harder level to achieve) was only worth 2.3 times a 3*. The inequity of this was acknowledged and the weightings were changed to 0/1/3/9 for subsequent funding years. This will have the effect of reducing the movement in funds to just 1%, down from the 3% trumpeted by HEFCE immediately after the 2009/10 funding allocations were announced.

The importance of reputation

Funding implications flow not just from the allocations received after PRFS assessments, but also as a result of the reputation achieved as a result of this process. Li *et al.* (2008, p. 6) point out that in a climate in which government funding for public universities is in decline in many countries, the continued viability and prosperity of a university can depend on its ability to attract external funding, which in turn can depend heavily on the institution's reputation. Success or failure validated by PRFS assessments can have significant flow-on effects outside the immediate funding allocations.

Reputation has gained increasing importance for Australian universities with the easing of restrictions on overseas students and the enrolment of full fee-paying students. In fact, international students are so important to Australia that education services are now Australia's third largest export, behind coal and iron ore (Baty, 2010). Many institutions appear to use the revenue they receive from these students to cross-subsidise the cost of research.

Reputation can be gained from sources other than PRFS. International rankings of universities, such as those published regularly by Jiao Tong University and the *THE*, have a strong influence on the regard in which institutions are held, and in turn have strong funding implications such as those mentioned in the preceding paragraphs. This situation exists even in the absence of any PRFS.

On human resource issues

For many years, concerns have been raised about the impact of PRFS on several personnel issues. Most relate to the impact of systems that require universities to select and submit for assessment only the work of "research-active" staff. The implications of this requirement are manifold. It is believed to affect the morale of staff who are not included in their institution's submission; to be biased against women, early career researchers and ethnic groups; to have serious implications for the autonomy of an individual's line of research; and to have led to an active "transfer market" for leading academics.

McNay's study of the RAE also highlighted more general changes in job descriptions, recruitment criteria, creation of posts, appraisal systems, career patterns and rewards:

> *"Previously, teaching in the PCFC sector and teaching and research in the UFC sector were conducted by permanent staff and a significant cadre of casual research workers. The greater emphasis on research has prompted some institutions to consider career paths for researchers, to provide bridging funds and to introduce more conscious management and appraisal of research staff. On the other hand, in some cases teaching posts have become casual to provide relief for permanent staff to do research. In general, a greater range of contract terms is being used."* (HEFCE, 1997, para. 81)

Hong Kong, China, reports similar responses. The focus on publications in the RAE has now been incorporated into the management practices of many universities. Most have established performance standards for academics which dictate the expected number of publications for promotions and appointments (Li *et al.*, 2008, p. 7).

The United Kingdom's RAE provides the most fertile ground for understanding the impact of PRFS on these issues, and studies undertaken on its impact provide some data to assist in distinguishing between perceptions (or possibilities) and reality. The two issues most often referred to are the designation of research-active staff and the transfer market in academic staff. These are discussed in more detail below.

Identifying research-active staff

A number of PRFS, such as those in the United Kingdom, Hong Kong (China) and New Zealand, require the identification of research-active staff. Only those with the requisite output in the specified time period are submitted, or have their work submitted, for assessment.

McNay found a significant impact on morale after the 1996 RAE when institutional submissions became more selective and only research-active staff were submitted. In the 1996 exercise, a survey of academics found that nearly two-thirds of departments excluded researchers from their submissions, primarily because of the quality of their research (Evaluation Associates Ltd, 1999, p. 23). A study by the Association of University Teachers found the selection of research-active staff biased against women; men were 1.6 times more likely to be entered in the RAE than their female counterparts (Baty, 2004).

Even when handled sensitively by management (and this was not always the case), selectivity became divisive and had a significant impact on staff morale and on the collegiality of institutions more generally (HEFCE, 1997, para. 82-86). It was characterised as the most "traumatic" effect of the RAE. HEFCE postulated that the increased stress on staff could "be mitigated by

greater institutional encouragement and support", though no studies have reported on whether such support mechanisms have been put in place (HEFCE, 1997, para. 100).

The need to identify research-active staff is also a concern in relation to New Zealand's PBRF, for similar reasons.

The academic transfer market

In the United Kingdom's RAE, universities are assessed on the basis of the publications of researchers currently on staff. Recently recruited academics bring with them all the outputs they produced in the preceding six years (the usual census period for assessment). Even if academics have only been on staff for a matter of months, their full track record for the period may be submitted for assessment by their current institution. This has led to what is commonly termed the transfer market in academics leading up to each assessment.

This transfer market was widely regarded as the "most newsworthy aspect of the exercise" in the run-up to the 1996 exercise and its existence was blamed on the RAE (HEFCE, 1997, para. 88-97). It is difficult to quantify the effect. McNay found that only 2% of staff in his survey had moved institutions in the 18 months prior to the 1996 RAE. However senior administrators commented that the 2% were all top-level staff (McNay, 1998, p. 20). In addition, the impact is not solely seen in movement between institutions. The HEFCE report notes that while movement between institutions in the lead-up to the 1996 exercise may not have been remarkable, nearly half the staff surveyed had been promoted in the previous five years. Conceivably, the 1996 RAE had a greater effect on internal movements than on external ones (HEFCE, 1997, para. 90).

This issue was a topic of considerable interest in the press in the lead-up to each subsequent exercise. In defence of its system of selectivity, HEFCE noted that even if the RAE had increased staff movement, this should not necessarily be construed as a negative consequence – it could be a just reward for the most talented researchers (HEFCE, 1997, p. 14).

On productivity

Increased publication output appears to be a common impact of PRFS, irrespective of the model used, and this has generated a great deal of attention. Much of the discussion is anecdotal, but it is the one impact on which there is considerable bibliometric analysis, accompanied by a belief that it is possible to demonstrate the causal effect of the assessment systems. The United Kingdom, Australia, Spain and Norway have been the focus of detailed studies.

The United Kingdom's RAE

A recent study, which examines annual data across the full period covered by all UK RAEs, looks more directly at the link between RAE cycles and changes in publication behaviour. It finds "pronounced 'timing effects' prior to a RAE" (Moed, 2008, p. 157). The data show:

- The United Kingdom's share of publications in the Web of Science increased significantly in the four years before the 1992 exercise in which assessment was based in part on total publication numbers.
- With the shift to a focus on quality in the 1996 exercise, the United Kingdom's share of publications in the Web of Science declined, but its share of higher-impact journals increased.
- The initial drop in productivity prior to the 1996 RAE appears to have lessened leading up to 2001 RAE as academics focused more on productivity and collaboration, perhaps in an effort to stimulate the number of staff judged research-active.

One note of caution applies to Moed's study: it used aggregate data for the United Kingdom and the analysis was not limited to higher education institutions. While universities account for the majority of research publications, a significant minority come from the hospital and non-profit medical research sectors. Ideally, separating the sectors could provide better information for testing the existence and strength of the universities' response.

Even though the focus of the 1996 RAE was quality, nearly two-thirds of respondents to the Evaluation Associates' (EA) survey agreed that it had encouraged them to publish more, and that the focus was predominantly on peer-reviewed journals (Evaluation Associates Ltd, 1999, p. 17). These responses are at odds with Moed's findings, which show that prior to the 1996 RAE productivity actually declined, with a stronger focus on quality than on quantity. It is an example of the mismatch between perception and reality, and raises questions about the extent to which anecdotal evidence, or survey responses, can provide an accurate picture of behavioural responses.

Australia

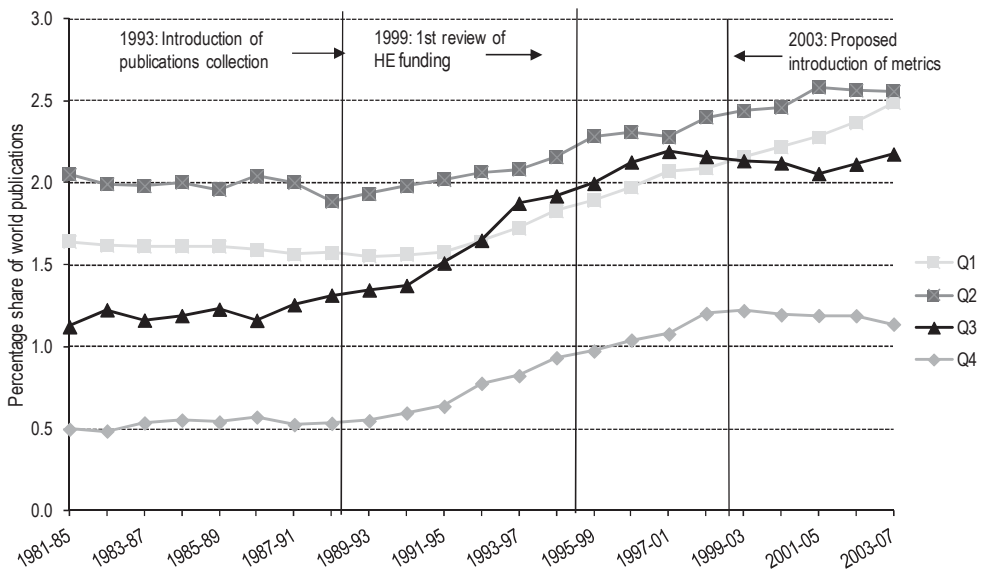
The Australian government has been using quantitative formulas to allocate the research component of the university block grant for nearly 20 years. A detailed analysis of the Web of Science data published in 2003 demonstrated a clear link between the introduction of this PRFS and an increase in productivity (Butler, 2003). By demonstrating that the identified trends applied to all fields of research, but were only present in the

university sector, and not in medical research institutes and other government research agencies, a direct causal relationship could be confidently implied.

Since that publication, the research policy landscape in Australia has shifted significantly. While the research block grants continue to be distributed on the same basis, a number of new initiatives have shifted the focus away from a purely quantity-driven calculation to quality. In 1999, a review of higher education funding first mooted the idea of dropping the element of the formulas based on simple publication counts. This did not happen immediately, but in 2003 a proposal was introduced to distribute funds on the basis of an RAE-style assessment system which would incorporate citation analysis where appropriate. A change of government saw this initiative fail, but a new initiative was developed which, while not yet linked to funding decisions, will assess the quality of research in Australian universities.

Figure 4.1 provides some indication of the response of universities to the various Australian policy initiatives.

Figure 4.1. Australian university responses to government funding initiatives



Note: Q1-4 indicate the quartile of journals, as measured by citation impact. Q1 is the highest quartile; Q4 is the lowest.

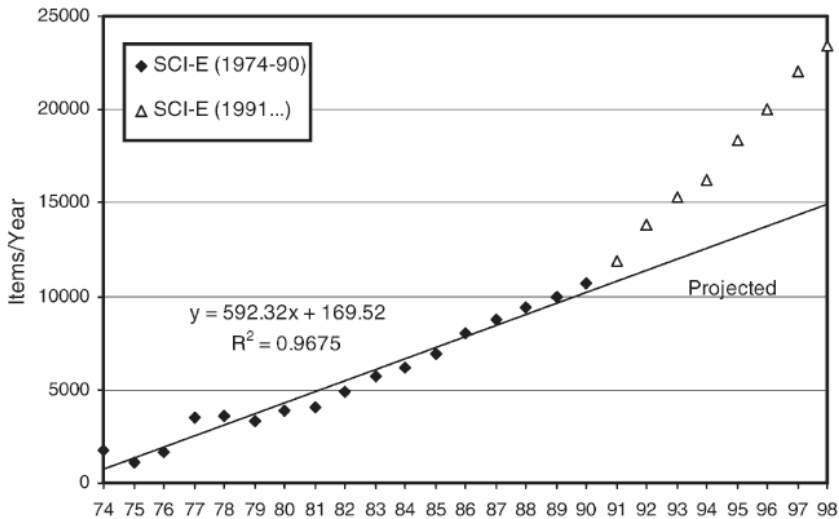
Source: Author.

The shaded area in Figure 4.1 reflects the analysis published in 2003, which clearly indicates the sudden increase in publication output from Australian universities after the introduction of the funding formulas, with a particularly sharp increase in the bottom two quartiles of journals (Q3 and Q4), as judged by citation impact. Another seven years of data have been added to that analysis and show that the trends in Q3 and Q4 flattened when the government's focus shifted to quality assessments, particularly after 2003.

Spain

In 1989, the Spanish government introduced an incentive system for researchers based on publication output. Refinements to the system in 1994 saw increased emphasis on the use of international journals indexed in the Web of Science. A strong link was found between these policy initiatives and a marked increase in output (Jiménez-Contreras *et al.*, 2003). As with the Australian study, many possible alternative explanations for the increase in productivity (such as increased international mobility, funding from non-government sources, etc.) were examined. None of the factors reviewed was able to explain the change in growth rates after 1989. This gave the authors confidence in establishing a causal relationship. Figure 4.2 shows the change in publication output over predicted trends after the PRFS was introduced.

Figure 4.2. Data-based predicted evolution of Spanish scientific production from 1974 to 1990



Note: SCI-E is SCI Expanded (a version of the SCI that contains more indexed journals than the CD-ROM version).

Source: Reprinted from Jiménez-Contreras *et al.* (2003), p. 131, with permission from Elsevier.

Spanish output in the Web of Science moved well above the long-term trend line after the introduction of the research incentive system in 1989. The authors believe that while the monetary incentive initially caused a surge in output, the direct monetary rewards were too small to explain fully the continued increased productivity. They believe the increased focus on research, and its associated influence on prospects for promotion, led to the continuing upward trend (Jiménez-Contreras *et al.,* 2003, p. 136).

Norway

In Norway there are concerns that the new PRFS might increase the number, but not necessarily the quality, of publications (Strehl, 2007, p. 51). However, since the introduction of the new system, there has been a substantial increase in publications, and output in both ranks of journals has increased at similar rates (Sivertsen, 2010, p. 26). There is some hesitancy in attributing this improvement to the new model, although Web of Science data show Norway's share of output increasing (2000-08), when that of other Nordic countries is decreasing (Sivertsen, 2009). The increase is in all journal quartiles. A number of qualitative changes have facilitated the increase in productivity – stronger institutional incentives for research; research seen as a communal as well as individual responsibility; increased internal awareness of publication activity; and improvement of research management practices (Sivertsen, 2010, p. 27). The Norwegian Ministry of Education saw the use of a two-level indicator as greatly reducing the risk of an explosion of publication activity at the expense of quality (European Commission, 2010, p. 121).

New Zealand

While the first PBRF did not take place until 2004, the proposal for a performance-based research funding system was first made in 2001. In most universities this triggered a marked increase in publication activity. When a production function approach was used to model the research process in New Zealand universities, the results corroborated the findings on increased productivity (Smart, 2009).

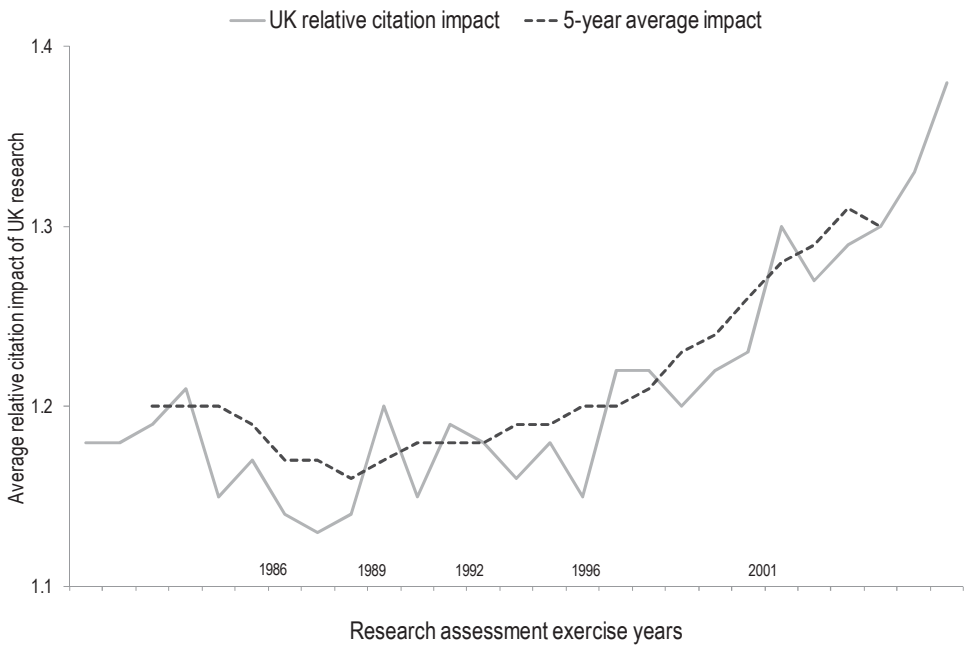
Flanders

The Flemish government explicitly acknowledged the danger of a focus on quantity at the expense of quality if their BOF key continued to rely on publication counts (Debackere and Glänzel, 2004, p. 272). It established a working group to look at refinements to the key that would incorporate citation impact measures. It has also moved to incorporate an indicator similar to the one used in Norway to include output from the social sciences and humanities (Sivertsen, 2010, p. 22).

On quality

A number of the studies described above also address the issue of the impact of PRFS on research quality. The Australian study clearly showed that, with the spotlight solely on productivity, there is a perceived drop-off in quality due to a greater tendency to publish in journals with less impact. With the change in focus to assessing quality, researchers appeared to maximise publication in journals with higher impact. There is no proof that, in itself, this denotes an improvement in the quality of research outputs, but a strong correlation is known to exist between publication in journals with high impact and quality of research, particularly at high levels of aggregation.

Figure 4.3. The relative citation impact of the UK research base, 1981-2007



Note: The average relative citation impact is citation counts normalised to world average by year of publication and journal category. The five-year average citation impact is the impact in a moving five-year window.

Source: Adams and Gurney (2010), p. 10. Reprinted from data and analysis by Thomson Reuters, published by HEPI (Higher Education Policy Institute) as "Funding Selectivity, Concentration and Excellence: How Good Is the UK's Research?"

As mentioned, in Norway, there were strong initial concerns that the system would increase the number, but not necessarily the quality, of publications (Strehl, 2007, p. 51). However, publication numbers in outlets classified in both the upper and lower ranks have increased at a similar rate (Schneider, 2009, p. 374).

There is a widely held belief that the RAE has been a driver of improved research quality in the United Kingdom (UK Government, 2006, para. 4.6). Bibliometric analyses confirm the improvement in UK science, in terms of relative citation impact. Most bibliometricians caution against asserting that a causal relationship has been proved, but the evidence that this is what the data reveal is mounting (Adams and Gurney, 2010, p. 3; Moed, 2008, p. 159), as Figure 4.3 demonstrates.

Figure 4.3 indicates the timing of the UK RAE cycles. The data show a downward trajectory in citation impact until the introduction of quality-based assessments in the 1992 RAE. The improvement in UK research impact has continued with each subsequent round of the RAE.

On teaching

A concern expressed in relation to all PRFS was the impact on the quality and standing of the teaching component of academics' work. In the United Kingdom, universities were required to report the total number of staff appointed to teaching-and-research and research-only positions, enabling calculation of the proportion of staff submitted for assessment. One way of improving this proportion was to move less research-active staff to teaching-only appointments.

One of the challenges faced by policy analysts is that while methods for assessing research quality are tried and tested, attempts to assess and reward teaching performance have proved much more problematic. As a result, discussions of the impact on teaching are largely anecdotal, with the exception of a research report from the UK Association of University Teachers that adds some evidential weight to the concerns. A few representative examples of these concerns are given below.

In terms of the effect of the RAE on teaching in the United Kingdom, research undertaken by the Association of University Teachers appears to substantiate claims that, in the lead-up to the 2008 exercise, less research-active staff were being moved to teaching-only contracts (Sanders, 2005). The belief was that this was a reaction to the need to select research-active staff for submission to the RAE and that it implied that teaching-only staff had lower status in the university system than staff with research activities.

When Ireland's new Recurrent Grant Allocation Model (RGAM) was phased in from 2006, there were concerns that it would divert resources from teaching to research, yet only 5% of the core budget is earmarked for research (Strehl, 2007, pp. 39, 50).

In Hong Kong, China, concerns that the RAE places "undue emphasis on research" given the "significant marginal returns" have been raised by the body responsible for distributing the funding (UGC, 2006, p. 4).

In New Zealand, there have always been concerns that any PBRF system will drive a wedge between the teaching and research functions of a university, yet studies that seek to determine whether or not this has occurred do not appear to have been carried out.

On discipline mix

The main concerns in relation to discipline mix centre on the real or perceived advantages or disadvantages that are thought to flow to particular discipline groups. Any quantitative assessment is generally seen to favour science, technology and medicine (STEM), at the expense of the humanities, social sciences and arts (HASS). This is primarily because accepted proxies for assessing research quality exist for STEM subjects in the form of citation analyses, but equivalent measures are not so readily available for the HASS disciplines. A comprehensive review of the issues faced by HASS disciplines is contained in a report published by the Australian Council for the Humanities, Arts and Social Sciences (CHASS, 2006).

In Norway too, where the PBRF does not incorporate any qualitative assessment, there was concern that areas of research in which the production of articles/monographs is not the norm will be discriminated against (Strehl, 2007, p. 51).

In the United Kingdom, concerns have focused more on disciplines with an emerging research culture than on a simple difference between STEM and HASS disciplines. In its response to a consultation on the RAE overseen by the joint UK funding bodies, the Department of Health raises its concerns about disciplines such as nursing and allied health professions:

> *"We have had mounting concerns that less well established disciplines such as nursing research or Allied Health Professions' (AHPs) work have been disadvantaged in a process where the quality standards have risen faster than such new subjects could be expected to keep up. This is discouraging at the personal level and financially damaging at the institutional one. This becomes even more pronounced since such new topics are frequently based in newer universities with only limited research resources."* (UK Department of Health, 2001, p. 4)

While the debate on the implications for the discipline mix within universities has not generally pitted STEM against HASS, there remain some concerns about the comparability of results across disciplines, particularly between the sciences and social sciences (Corbyn, 2008). These concerns exist even within these two subject groups and do not appear to have been alleviated by changes to the panel structure for the 2008 RAE that was designed with this issue in mind.

On the focus of research

A general concern of all PRFSs, whether based on qualitative or quantitative assessment, is that they favour "mainstream", disciplinary-based, basic, "safe" research at the expense of applied, interdisciplinary or speculative research. These concerns are regularly voiced in the academic press, yet little research has been undertaken to prove or disprove these biases. As Gläser notes succinctly:

> *"Our knowledge about the development and effects of RES [research evaluation system] has developed rather unevenly. While the science policy processes and organisational dynamics have been investigated for several countries, we have just begun to explore how RES interact with scientific communities and their research."* (Gläser, 2008, p. 246)

In Spain, concerns about the Evaluation of Research Activity policy span a number of related themes, as it is believed to lead researchers to focus on research that is attractive to Web of Science journal editors, and there is a fear that more applied research, and research with a local/regional focus, may suffer (European Commission, 2010, p. 123). However, studies to determine whether these fears have been realised do not appear to have been undertaken.

Evidence gleaned from the literature on a number of themes relating to the focus on research is presented below.

Interdisciplinarity

A major concern for PRFS, based either on field-based quantitative indicators or peer assessments by panels convened along disciplinary lines, is the impact on interdisciplinary work. It is extremely difficult to find clear evidence on whether such systems are indeed biased against such research.

Quite often perceptions, either from survey responses or discussed on an anecdotal basis in the press, can be quite different from the reality. An example is the regularly discussed concern on the bias of discipline-based systems against interdisciplinary research. The view that the RAE was biased against interdisciplinary research is widespread, and accepted within

government circles, and was one of the reasons quoted for moving to a REF (UK Government, 2006, para. 4.7). It was also a major motivation for the Roberts Review of the RAE after the 2001 exercise (Roberts, 2003, p. 22).

Yet two studies on this issue, undertaken before these policy changes were mooted, found no evidence to support the assertion. McNay's study found that while the number of staff who reported that they had moved away from interdisciplinary work was relatively small, almost half those surveyed felt that the RAE hindered interdisciplinary work (McNay, 1998, p. 20). The Evaluation Associates study found that while nearly a quarter of researchers believed that the RAE was a strong inhibitor of interdisciplinary research, the ratings achieved by departments with a high proportion of interdisciplinary researchers show that there was in fact no such discrimination (Evaluation Associates Ltd, 1999, pp. 13, 28).

McNay's 1997 survey demonstrates that reactions can vary by type of institution and that general statements on an issue may not provide an accurate picture. While a significant minority of researchers, and nearly half of research managers, felt it was not a good strategy to pursue inter-disciplinary work, many staff in the post-1992 universities focused on applied/interdisciplinary work, which they saw as their strength (McNay, 1998, p. 20).

"Blue skies" research

McNay's survey found that nearly half the research managers surveyed felt the RAE hindered the pursuit of new research areas or risky "blue skies" research (McNay, 1998, p. 20). In Norway, the principal fear regarding the new PRFS appears to be that "mainstream" research will be advantaged (Strehl, 2007, p. 51). The Hong Kong, China, RAE was also perceived to have a "narrow focus on traditional research" (UGC, 2006, p. 4).

The RAE time scales compounded these concerns for staff and a significant minority (around one-quarter) reported avoiding new lines of research and speculative topics because they believed quality outputs could not be achieved by RAE submission deadlines (McNay, 1998, p. 20). This was corroborated by Evaluation Associates' study of the 1996 RAE, which found that researchers felt pressure to get results published early to meet cut-off points for the exercise (63%), and a small minority felt that there was little incentive to focus on longer-term research (Evaluation Associates Ltd, 1999, p. 17).

Research on the periphery

A third theme in this area relates to smaller nations and/or those seen to be on the periphery of the large centres of research in the United States and the United Kingdom. University rankings are believed to lead to a neglect of local or regional issues in order to publish in high-impact international journals (Hazelkorn, 2009, p. 2). This is particularly true of PRFS that specifically aim to increase publication in the international journals, as happens in Spain (Jiménez-Contreras *et al.,* 2003). There are legitimate concerns that national journals and research priorities will be neglected in the rush to publish in Web of Science journals. PRFS funding systems that incorporate ranked outlets and/or citation measures have the potential to encourage such a response.

Fifteen years ago Arunachalam raised concerns about the mismatch between the topics on which Indian medical researchers published and the areas of research related to the country's most pressing medical problems (Arunachalam, 1995). While this did not result from PRFS but from academic incentives that rewarded researchers with career advancement for a strong international publication record, it demonstrates that concerns about a reliance on databases with country/language biases are real.

Applied research

A fourth theme focuses on concerns that research targeted at policy, social interventions, professional practice, etc., will be disadvantaged. This is of particular concern in PRFS systems based primarily on quantitative indicators, as much of this research is published in the "grey literature" that falls outside the ambit of standard performance indicators.

Researcher autonomy

One issue that has gained prominence recently relates to concerns over researcher autonomy. With the increased incidence of departmental micro-level research management strategies, there is concern that researchers are being pressured to stay within their own field of expertise and not move down new paths. A quote from one of the respondents to the survey describes these pressures:

> *"Unfortunately the RAE categories and decisions are, at least in my institution, permeating decisions about research activity. They're increasingly looking for RAE publications; in a sense, monitoring people's research and parts of their careers."* (RIN, 2009, p. 36)

Nearly one-third of staff surveyed by McNay felt that the direction of their research was increasingly conditioned by the collective priorities of their group or department, rather than their individual preferences (McNay, 1998, p. 20).

In Norway there are also concerns about a loss of autonomy. "Research is now perceived as a common and institutional responsibility, not only as an individual task." (Sivertsen, 2009, slide 38)

On collaboration

Concerns about the effect of PRFS on collaboration appear to relate to specific details of the assessment methodology. The choice is generally between whole counts, *i.e.* giving each institution (or author) full credit for an article on which their name appears (which leads to double counting) and fractional counts, *i.e.* an article only counts as a single unit and each institution (or author) receives a fraction of the credit.

Concerns about the effect of fractional counts on collaborative activities led Australian governments to reject this methodology in favour of whole counts. Norwegians appear less concerned and believe their use of fractional counts has not resulted in a decline in collaborative activities (Schneider, 2009, p. 372). Schneider believes that "'invisible colleges' and social networks within research specialties have eventually ensured collaboration". It is believed that the dependence of research on collaboration will counteract any adverse behaviour that might result from the funding model.

In contrast, in response to a survey on the 1996 RAE, researchers felt that the RAE did not encourage them to work more collaboratively, particularly with researchers or users outside the university sector (Evaluation Associates Ltd, 1999, p. 50).

Countries such as Australia and the United Kingdom continue to have dual funding systems and specific schemes aimed at stimulating collaborative activities. Many EU funding programmes require extensive collaborative networks. These work to soften the impact of any system that might otherwise seem to work against this trend.

On institutional management practices

The first response to the introduction of a PRFS often comes from senior management – it takes time for individual academics to feel its effects. The implications of any PRFS do not appear on an academic's radar until specific institutional policies put it there, *e.g.* the need to provide information; individual funding incentives for individual researchers to improve performance; changes to requirements for promotion; etc.

Information management systems

A number of countries report significant improvement in the information management systems of universities. With the increased need for data, at ever increasing levels of detail, many universities have redesigned their information management systems. One of the main drivers is the need to bring together data from a range of different internal sources: data on students, staff, competitive grants, external income, research publications, etc.

The rise of PRFS and the increased visibility of university ranking tables have led universities, among other things, to a "renewed emphasis on the accuracy/amount of data gathered and shared with third parties" (Hazelkorn, 2007, p. 13).

Research management also is claimed to have improved in Norway with the aid of complete bibliometric information about research activities (Sivertsen, 2009, slide 38). The same is also regularly claimed for Australia and the United Kingdom.

Institutional repositories

In both Australia and the United Kingdom, moves to introduce a metrics-based system of assessment has led universities to expend considerable effort on improving or establishing institutional repositories for housing publications that are likely to be included in the assessment. This is seen to be a particularly beneficial, unintended consequence of the introduction of a PRFS. Australian librarians noted:

> *"The imminent introduction of the RQF [Research Quality Framework] has served to justify, and hence to hasten, the introduction of a repository in some universities, while encouraging better communication between the research office and repository managers."* (Henty, 2007)

This is a trend that is also gathering pace in the United Kingdom.

Research management

When first introduced, PRFS can have a profound effect on the focus and styles of research management within institutions.

A survey of universities after the 1992 UK RAE found that it had become a driver of institutional research planning, particularly when it became clear that the assessment exercise would be conducted regularly (HEFCE, 1997). However, it was not always clear that these changes could be linked directly to the RAE. A significant minority of respondents (close to one-third) also pointed to the policies of other government agencies, research councils and funders as important drivers of improved, strategic

research management. McNay found a greater focus on research in the post-1992 universities and colleges (McNay, 1998).

The view that these changes could be linked, at least in part, to the rise of assessments/rankings was supported by a more recent survey. Respondents reported that new organisational sections were established, or individuals assigned, "to deal with indicator improvements and monitor rankings" (Hazelkorn, 2007, p. 13).

Another recent study made a related point: many changes may occur anyway, but assessments/rankings made these happen quicker (CHERI and Hobsons Research, 2008, p. 34). The rise of university rankings is also believed to lead to strong reactions from senior management.

Departmental restructuring

One consequence of research assessment that has been directly linked to the RAE has been the internal restructuring of universities after the results of the latest exercise are released.

In some cases, universities respond to poor RAE results by closing departments. After the most recent RAE in 2008, the University of Liverpool announced the likely closure of three departments – statistics, philosophy and politics, and communications studies. All were deemed by the RAE assessors to have no "world-leading" (4*) activity (Newman, 2009a). The university is also contemplating closing other departments that were in the lowest quartile (based on 3* + 4* outputs) – civil engineering, cancer studies, dentistry, American studies and sociology. This, in spite of the fact that the University of Liverpool was an overall "winner" in RAE 2008 (its QR funds increased by 21.3% – GBP 4.4 million). It was also contemplating providing no support for staff assessed as 1*.

It is feared the poor showing of politics in the 2008 RAE (rated 62 overall out of 67 disciplines, based on the average research grade) will lead to a number of department closures, even though the RAE results were at odds with an Economic and Social Research Council review which assessed the discipline's overall standing as high (Newman, 2009b). This demonstrates the RAE's power to drive management strategies in the UK higher education system.

In France, some institutions use the AERES exercise as a "benchmarking exercise" and try to "improve their rating by closing down sectors which show a poor performance" (European Commission, 2010, p. 97).

Strategic recruitment

Some of the implications of the selective submission of research for assessment, particularly for individual researchers, are discussed above. This section looks more specifically at the strategies that institutions have employed in deciding the proportion of staff to submit for assessment.

The issue is perhaps most clearly illustrated in relation to the most recent RAE. A number of 2008 RAE panels (*e.g.* economics and econometrics; accounting and finance; business and management studies) believed that some universities excluded research-active staff in order to gain a higher ranking, while others drafted in "research stars on 'unusual' contracts who were not fully integrated members of the research team" (Newman, 2009c). This occurred despite the chair of the social sciences panel having warned, well over a year earlier, that he would be looking at the possibility of "game playing" and urging universities to enter all their research-active staff (Lipsett and Tysome, 2006). However, it was difficult to determine the extent of this practice empirically because there was no information on the proportion of staff submitted owing to ambiguous definitions of eligibility.

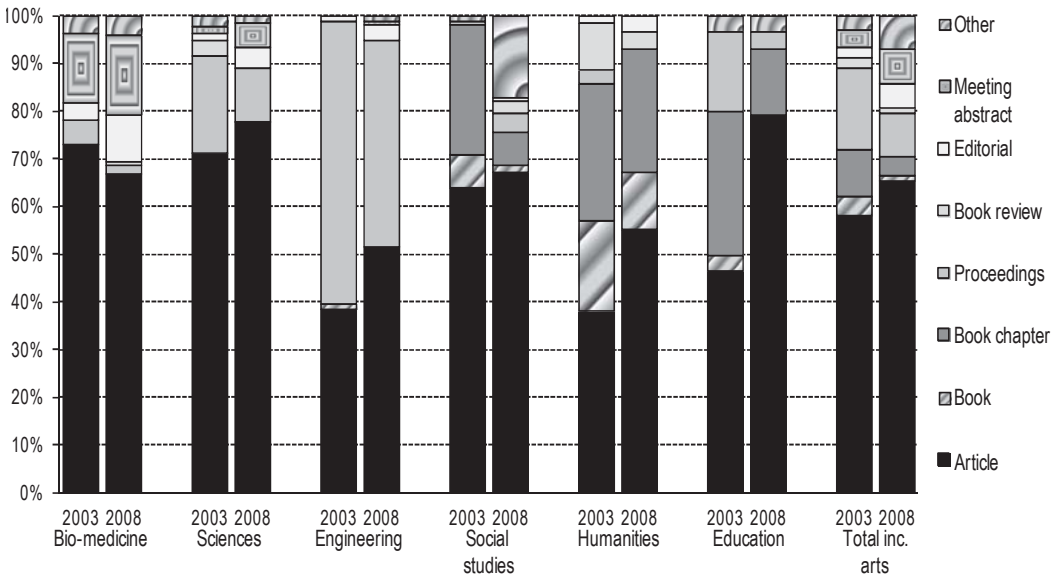
HEFCE had recognised the incentive to omit staff from submissions in order to improve their ranking, and the new system of distributional scoring was introduced in part to overcome such practices. It does not appear to have been successful (MacLeod, 2006). Australia hoped to lessen some of the game playing by including *all* research. However it followed the United Kingdom's lead in using a census date approach to determining the institutional affiliation of researchers, so it is expected that some form of transfer market will arise (or already has arisen).

On where to publish

One of the major concerns about PRFS relates to the possibility of publication practices being altered to suit what is seen as the best strategy for the assessment methodology. In particular, there is concern about the rise in the use of journals in disciplines that have traditionally favoured book or conference outlets.

A recent online survey of over 800 UK academics provides some very clear evidence on changes that are occurring in academic publishing patterns and on the reasons why. Figure 4.4 provides clear insights into changes that have occurred in the last decade.

Figure 4.4. Change in publication outlets by major field of research, 2003 to 2008



Source: (RIN, 2009, p. 16).

Concerns about these changes are encapsulated in the views of one of the respondents in that study:

> *"Sadly, I find myself increasingly moving away from publishing in journals which are important and read by a lot of colleagues, to publishing in high status journals instead. This had led to much longer delays [and] thus adversely affects science, but I feel the pressure to do this in order to advance career-wise." [Medical and biological sciences]* (RIN, 2009).

McNay's earlier 1997 study also found that most heads of research units (84%), and a majority of staff (55%), felt that the RAE had led to the targeting of key journals (McNay, 1998, p. 20), and many of those surveyed believed the RAE had led to premature publication (30% of heads of units believe it, 25% of staff admit to it).

As mentioned earlier, in addition to a changing use of publication outlets, the study found that work is being fast-tracked for publication and may be less thoroughly researched than in the past. Some analysts blame the "publish or perish" culture and the pressures this entails as the reason for the increased rate of retractions in scientific journals (Corbyn, 2010). Perhaps these two trends are linked.

On author behaviour

Multiple authorship has become commonplace in research publications, reflecting an increase in collaboration both within and between institutions and countries. However disciplines, and even different groups within disciplines, vary considerably in their standard traditions for author attribution. From an online survey of UK academics, a report conducted for the Research Information Network (RIN) determined the different practices among disciplines from over 800 responses (RIN, 2009). It is worth reproducing the summary of their findings here:

> *"The listing of authors in order of contribution (with first author providing the greatest contribution) is the most frequent practice in most disciplines except for the humanities where alphabetical order is the norm. But it is notable that in physical sciences, mathematics and social sciences alphabetical ordering and ordering by contribution are almost equally common. Notable also are the differences of practice within discipline groups. In medical and biological sciences, in physical sciences and maths, and in engineering and computing, ordering by contribution may frequently be modified by placing the senior researcher or grant-holder last. And while the most common practice with papers arising from research undertaken by students is to place the student first in the author list, a significant minority in medical and biological sciences and in social sciences place the supervisor first."* (RIN 2009, p. 26)

The implications of this for any performance-based funding system that is predicated, even in part, on quantitative measures related to publications is two-fold:

1. Any attempt to fractionate publications across collaborating authors/ institutions is fraught with difficulties, particularly if an attempt is made to weight shares according to level of contribution. Practices vary too widely to allow this to be done automatically. Until such time as authors routinely ascribe percentage contributions, it is unlikely that fractionation can be used without seriously disadvantaging some authors and/or disciplines.

2. Undertaking any analysis on the basis of first author only (as in the current Swedish model) is likely to have serious consequences, particularly at lower levels of aggregation. At the national and institutional level, the assessment of performance is likely to be accurate for most countries, but as soon as the methodology is used at lower levels of aggregations, serious distortions can arise.

Additional comments

Overt game playing

The strategies adopted by universities in response to PRFS are unsurprising. All seek to maximise their returns from a system with direct implications for the funding they receive. It can be difficult to draw the line between acceptable and unacceptable strategies. Much criticism has surrounded the "transfer market" in "star" researchers that occurs before each RAE round. But this can also be seen as a just reward for high achievers, something that was missing prior to the introduction of the RAE.

Yet undoubtedly some strategies are clearly against the interests of the sector. One unambiguous example relates to the new area of assessment based on webmetrics. Some universities have started to use link farms[1] and paid backlinks[2] to improve their positions (European Commission, 2010, p. 132). This clearly undermines the rankings, and the rankings developers have removed these institutions from their tables (Prieto Valverde and Fernández Pérez, n.d.). They have also discovered some universities that are hosting large numbers of academic papers authored by scientists that do not belong to those institutions.

A few instances of game playing have been identified in relation to the Australian system, which is based on publication counts. Quite detailed and restrictive definitions are applied to the four types of publications that can be included in a university's return – conferences must be "international" and journals must have an editorial board that goes well beyond a single institution. One university overcame the first issue by establishing a conference with "international" in the title and ensuring at least one participant from overseas, but 90% of the presentations were from the institution. An entrepreneurial academic overcame the restrictions on editorial board by establishing one consisting solely of his former PhD students, a number of whom had moved back to their former institutions. Both strategies were well outside the spirit of the process, but fortunately such extreme examples appear to be rare.

Perception and reality

The Research Information Network report clearly shows that researchers submit different kinds of outputs to those they publish – the sciences and engineering submit a much lower proportion of conference papers than in their overall publishing patterns; and the humanities submit a larger proportion of books (RIN, 2009). The survey of academics reveals that a quarter of those who responded believed that the RAE excludes some types of research output which they consider important. This was despite all

panels clearly stating that virtually all types of output are admissible and the guidelines clearly specifying that all should be treated equally. Clearly some academics receive misleading advice, either from institutional research managers or their colleagues.

In relation to peer review, a British Academy report put forward the argument that the criticisms directed at peer review related to "deficiencies of practice rather than the principle of peer review" (British Academy, 2007). The system itself was considered the best, but sometimes less than optimal in practice.

Inconsistent outcomes

Institutions do not always respond in identical ways. A poor assessment may lead one university to close down a department, but another may decide to invest more resources in that department in an attempt to improve outcomes in the next assessment. HEFCE noted that the RAE was said to have stimulated some institutions to recruit new or young staff with research potential, but also to have caused other institutions to adopt a more conservative approach of recruiting proven researchers (HEFCE, 1997, para. 101).

Responses may change over time. In the United Kingdom, institutions appear to have become more selective of the staff/publications/units they submit and more aggressive in their recruitment of "stars" in each successive RAE cycle.

Also, the problem of grade inflation in the 2001 RAE resulted in departments that had maintained a 5* rating losing funding while, at the same time, some departments that improved their ranking did not receive funding because departments on the lower ratings (1, 2, and 3b) were no longer funded (Tapper and Salter, 2003, p. 19).

Other influences

PRFS do not operate in isolation. Countries such as Australia and the United Kingdom operate dual funding systems, with significant amounts of money distributed through a number of research councils. The methods other funders use to determine the success of grant applications can soften the impact of PRFS, particularly if different assessment processes are used. In Australia, the Research Council uses a traditional peer review approach, in stark contrast to the blunt funding formula used to distribute the research block grant. In the United Kingdom, the systems are less dissimilar, and this is likely to reinforce the impacts of the PRFS.

Unrelated government policies can soften potentially negative unintended consequences. A number of countries have programmes aimed at facilitating cross-institutional collaboration, which work to overcome a potential bias against collaborative work in an assessment system that uses fractional counts (*e.g.* Sweden). In Australia, the existence of a significant pool of money to fund grants that link researchers with industry helps to counter a potential bias against more applied research.

The Evaluation Associates' study also listed a small sample of other factors that can influence research performance (and in particular interdisciplinary research). Teaching loads; the availability of equipment, software and data; contract funding; institutional "environment"; and the policies of research councils were some of the examples provided (Evaluation Associates Ltd, 1999, p. 17).

Responses to specific assessment measures

Many responses are generic, whatever the methodology employed. But some are quite specific to the particular methodology used. This was touched on in relation to the choice between fractional and whole counts in quantitative-based assessments. Some other measure-specific concerns are briefly covered here.

The recent RIN report reinforces the fears of many academics (22%, or nearly a quarter, of survey respondents) who believed the use of bibliometrics in research assessment would lead to game playing in the form of citation clubs and increased self-citation and citation of collaborators, although few indicated they would cite their rivals' work less often (RIN, 2009, p. 7). What all commentators on the effects of bibliometrics seem to forget is the role that journal editors and reviewers play as gatekeepers of appropriate behaviours.

Assessing only a specified proportion of total output can result in misleading assessments of total research. One panel chair feared that the selectivity used in some institutions could lead to very poor results in disciplines for which the perceived wisdom was that UK research was very good (Corbyn, 2008).

Responses from institutions may depend on details of the process. There are concerns in the United Kingdom that moves to force universities to submit all research-active staff may lead to even more destructive game playing than the existing selective process which encourages the poaching of "stars" (Roberts, 2003, p. 52), but in Australia the inclusion of all staff (and all outputs) does not appear to have led to a major shift in the classification of staff.

Broader impact

In some instances, the introduction of PRFS can have an impact beyond the higher education or research sectors on which the assessment focuses. Research is not an isolated endeavour. It involves a myriad of inter-linked stakeholders, all with an interest in the processes used in those assessments, and the implications, if any, for them. Some examples of the broader impact of assessment systems are described below.

In Spain there is a perception that several Spanish journals have improved their manuscript assessment and peer review processes in an attempt to be indexed in the Web of Science (European Commission, 2010, p. 123). The scientific publishing industry has a keen interest in the mechanics of these systems as they can have a direct bearing on the attractiveness of their journals as outlets for the work of researchers. Journal editors and publishers took a very keen and active interest in the Australian exercise to classify over 20 000 peer-reviewed journals into four quality bands. The work done by committees attempting a similar exercise for the European Reference Index for the Humanities under the aegis of the European Science Foundation also gained their keen attention.

As with improvements in the standards of Spanish journals, PRFS can have other external benefits. Public pressure can lead to improvements in the quality of data, as in the development of the Berlin Protocol for university rankings. It also occurred with the German CHE rankings, which initially only included professors' publications. There was a strong belief that this distorted the outcomes, and the collection of data was subsequently expanded to include the publication of all staff holding a doctorate (European Commission, 2010, p. 104).

Even the rise of university rankings can be seen as an indirect result of the introduction of PRFS in a limited range of countries – they fill the gap for countries in which no such system exists (European Commission, 2010, p. 111). The developers of the Jiao Tong rankings in Shanghai have often said that the prime motivation of the construction of their index was to provide a method, however crude in its initial form, to answer the question: "Does China have any world class universities?"

Conclusion

Hicks (see Chapter 1) lists 12 countries with performance-based funding systems, yet there is a relative scarcity of evidence-based analysis of the impact of these systems in the literature. It is possible that some may be buried in the grey literature of government reports. While Norway is known to be very active in this area, and researchers have presented the results of a number of studies at recent conferences, it is difficult to find detailed reports. It is likely that a number of studies will only be written in the language of the relevant country, as the primary audience will be local government agencies. This may explain a lack of studies from Poland and the Slovak Republic.

It is likely that considerable information could be unearthed, and it is hoped that, as is occurring in Italy, scholars will start to scrutinise these systems and publish more evidence-based assessments of their impact. They are certainly needed. Without hard evidence, anecdotes will hold sway.

This literature review shows the mismatch that can occur between perception and reality. It raises questions about the extent to which anecdotal evidence, or survey responses, can provide an accurate picture of behavioural responses. For every anecdote about a particular response to a PRFS, it will nearly always be possible to discover an anecdote that suggests the contrary. Similar contradictions in responses to survey questions have also been highlighted.

Yet the distribution of research funds on the basis of the assessment of performance is here to stay. The architects of PRFS may well anticipate, and even encourage, some of the more obvious impacts of the schemes they introduce. But academics are very creative – they can and will respond to such measures in novel and unforeseen ways. And the institutions they serve will always seek to maximise returns, as is their responsibility. Many impacts occur, no matter what method is employed. Redesigning the system will therefore not change anything – what is needed are alternative policies to overcome unwanted responses from academia.

This leads back to one of the central points made in Whitley and Gläser (2008): What needs to be studied is clear. The challenge is to facilitate the prioritisation of studies on these issues. Some potential studies are easy to identify and scope. They include:

- A broad, multinational assessment of publication trends that compares sectors/countries subject to PRFS with those that are not, thus providing a clearer picture of causality.
- Bibliometric analyses of trends in co-citation and co-authorship patterns to gain insight into the effect of PRFS on interdisciplinarity and collaboration.
- Disaggregating data analyses down to disciplines to investigate differential responses to PRFS.
- Analysing trends in publication strategies (perhaps expanding the RIN study to other countries).
- A detailed study of staff data to determine changes in the classification of staff and movement between institutions to provide some insights into questions such as gender bias, movement to casual and/or teaching only appointments, and the extent of the "transfer market".

For these studies the data are readily available, but they need to be co-ordinated and have a wider remit than a single country. It is more difficult to delve deeply into issues that require the input of sociologists of science and their qualitative assessment techniques: how the content of research has changed; whether applied research is suffering; what signals researchers are receiving and how they respond; etc.

What is clearly needed is a co-ordinated approach to the study of the impact of PRFS on all aspects of the academic endeavour, incorporating both quantitative and qualitative approaches. Without valid evidence on their impact, it is impossible to assess the efficacy of their use and develop strategies to overcome adverse consequences.

Acronyms

AERES	*Agence d'Evaluation de la Recherche et de l'Enseignement Supérieur* (France)
AHP	Allied Health Professions (United Kingdom)
ARC	Australian Research Council
BOF	*Bijzonder Onderzoeksfonds* (Flanders)
CHASS	Council for the Humanities, Arts and Social Sciences (Australia)
CHE	Centre for Higher Education Development (Germany)
CHERI	Centre for Higher Education Research and Information (United Kingdom)
DfES	Department for Education and Skills (United Kingdom)
ERA	Excellence in Research for Australia
HASS	Humanities, Social Sciences and Arts (United Kingdom and Australia)
HEFCE	Higher Education Funding Council of England
ITP	Institutes of Technology and Polytechnics (New Zealand)
PBRF	Performance-Based Research Fund (New Zealand)
PCFC	Polytechnics and Colleges Funding Council (United Kingdom)
PRFS	Performance-based Research Funding System
RAE	Research Assessment Exercise (United Kingdom)
REPP	Research Evaluation and Policy Project (Australia)
RES	Research Evaluation Systems
RIN	Research Information Network (United Kingdom)
RQF	Research Quality Framework (Australia)
STEM	Science, Technology and Medicine (United Kingdom)
THE	Times Higher Education
UGC	University Grants Committee (Hong Kong, China)

Notes

1. According to Wikipedia a link farm is: "any group of web sites that all hyperlink to every other site in the group".
2. The Wikipedia definition of backlinks is: "incoming links to a website or web page".

References

Adams, J. and K. Gurney (2010), "Funding Selectivity, Concentration and Excellence – How Good Is the UK's Research?", Higher Education Policy Institute (HEPI), *www.hepi.ac.uk/455-1793/Funding-selectivity,-concentration-and-excellence---how-good-is-the-UK%27s-research.html*.

ARC (Australian Research Council) (2009), *ERA 2010 Submission Guidelines*, *www.arc.gov.au/era/key_docs10.htm* (accessed 23 March 2010).

Arunachalam, S. (1995), "Research As If Relevance Mattered: Medical Research in India As Reflected by SCI 1981-85", in *Proceedings of the Fifth Biennial Conference of the International Society for Scientometrics and Informetrics*, ISSI, River Forest, IL, pp. 33-42.

Baty, P. (2004), "Sex Bias Limits Women in RAE", *THES*, 16 July, *www.timeshighereducation.co.uk/story.asp?storyCode=190047&sectioncode=26* (accessed 6 March 2010).

Baty, P. (2010), "Exposed and Vulnerable", *THES*, 22 April, *www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=411285&c=2*.

British Academy (2007), *Peer Review: The Challenges for the Humanities and Social Sciences*, *www.britac.ac.uk/policy/peer-review.cfm*.

Butler, L. (2003), "Explaining Australia's Increased Share of ISI Publications - The Effects of a Funding Formula based on publication Counts", *Research Policy*, 32(1), pp. 143-155.

CHASS (2006), "Measures of Quality and Impact of Publicly Funded Research in the Humanities, Arts and Social Sciences", *www.chass.org.au/papers/PAP20051101JP.php*.

CHERI and Hobsons Research (2008), "Counting What is Measured or Measuring What Counts? League Tables and Their Impact on Higher Education Institutions in England", issues paper, report to HEFCE, April 2008/14.

Corbyn, Z. (2008), "RAE Results May Not Reflect True Quality of UK Research, Warns Chair", *THES*, 4 December.

Corbyn, Z. (2009), "Reversal of Fortunes", *THES*, 5 March.

Corbyn, Z. (2010), "Retractions Up Tenfold", *Times Higher Education*, 22 March, *www.timeshighereducation.co.uk/story.asp?storycode=407838*.

Debackere, K. and W. Glänzel (2004), "Using a Bibliometric Approach to Support Research Policy Making: The Case of the Flemish BOF key", *Scientometrics*, 59(2), pp. 253–276.

DfES (2006), "Reform of Higher Education Research Assessment and Funding: Summary of Responses", *www.dcsf.gov.uk/consultations/downloadableDocs/RAE%20response%20summary %20250107.doc*.

Elzinga, A. (2009), "Evidence-based Science Policy and the Systematic Miscounting of Performance in the Humanities", *The Humaniora/Society Blog*, *http://humaniorasociety.wordpress.com/2009/04/28/evidence-based-science-policy-and-the-systematic-miscounting-of-performance-in-the-humanities/*.

European Commission (2010), *Assessing Europe's University-Based Research*, European Commission, Brussels.

Evaluation Associates Ltd (1999), *Interdisciplinary Research and the Research Assessment Exercise*, HEFCE, *www.rae.ac.uk/2001/Pubs/1_99/.*

Frölich, N. (2008), *The Politics of Steering by Numbers: Debating Performance-based Funding in Europe*, NIFU STEP, Oslo.

Gläser, J. (2008), "The Social Orders of Research Evaluation Systems", in R. Whitley and J. Gläser (eds.), *The Changing Governance of the Sciences*, Sociology of the Sciences Yearbook, Springer, Dordrecht, pp. 245–266.

Gläser, J. *et al.* (2002), "Impact of Evaluation-based Funding on the Production of Scientific Knowledge: What to Worry About, and How to Find Out", *Expertise for the German Ministry for Education and Research*, Vol. 31.

Hazelkorn, E. (2007), "The Impact of League Tables and Ranking Systems on higher education decision making", *Higher Education Management and Policy*, Vol. 19(2), p. 25.

Hazelkorn, E. (2009), *Pros and Cons of Research Assessment: Lessons from Rankings*, Dublin Institute of Technology, Directorate of Research and Enterprise, Dublin.

HEFCE (Higher Education Funding Council of England) (1997), *The Impact of the 1992 Research Assessment Exercise on Higher Education Institutions in England*, *www.hefce.ac.uk/pubs/hefce/1997/m6_97.htm*.

Henty, M. (2007), "Ten Major Issues in Providing a Repository Service in Australian Universities", *D-Lib Magazine*, Vol. 13(5/6), *www.dlib.org/dlib/may07/henty/05henty.html*.

Jiménez-Contreras, E., F. Anegon and E. Lopez-Cozar (2003), "The Evolution of Research Activity in Spain – The Impact of the National Commission for the Evaluation of Research Activity (CNEAI)", *Research Policy*, Vol. 32(1), pp. 123-142.

Li, B., J. Millwater and P. Hudson (2008), "Building Research Capacity: Changing Roles of Universities and Academics", paper presented at the Australian Association for Research in Education (AARE) 2008 International Education Research Conference, Brisbane.

Lipsett, A. and T. Tysome (2006), "A Nod to New Blood and a Warning to Game Players", *THES*, 3 February.

MacLeod, D. (2006), "Made to Measure", *The Guardian*, 31 January.

McNay, I. (1998), "The Research Assessment Exercise (RAE) and After: 'You Never Know How It Will All Turn Out'", *Perspectives: Policy and Practice in Higher Education*, Vol. 2(1), p. 19.

Ministry of Education, New Zealand (2008), "How the PBRF Has Shifted Research Funding", *www.educationcounts.govt.nz/__data/assets/pdf_file/0011/18848/Funding_factsheet.pdf*.

Moed, H.F. (2008), "UK Research Assessment Exercises: Informed Judgments on Research Quality or Quantity?", *Scientometrics*, Vol. 74(1), pp. 153-161.

New Zealand, Tertiary Education Commission (2010), Performance-Based Research Fund – Tertiary Education Commission, *www.tec.govt.nz/Funding/Fund-finder/Performance-Based-Research-Fund-PBRF-/*.

Newman, M. (2009a), "Liverpool May Drop Subjects Ranked Below 4* in RAE", *THES*, 12 March.

Newman, M. (2009b), "RAE Panel Hurt Field by Labelling It 'Weak'", *THES*, 26 March.

Newman, M. (2009c), "Reviewers Raise Concerns about RAE Game Playing", *THES*, 8 January.

Prieto Valverde, J.A. and M. Fernández Pérez (2010), "World Universities' Ranking on the Web 2010", *www.webometrics.info*/.

RIN (Research Information Network) (2009), *Communicating Knowledge: How and Why Researchers Publish and Disseminate Their Findings*, September, *www.rin.ac.uk/communicating-knowledge*.

Roberts, G. (2003), *Review of Research Assessment*, Report to the UK funding bodies, *www.ra-review.ac.uk/reports/roberts.asp*.

Sanders, C. (2005), "20% of Staff Now Just Teach", *THES*, *www.timeshighereducation.co.uk /story.asp?storyCode=196862&sectioncode=26*.

Schneider, J.W. (2009), "An Outline of the Bibliometric Indicator Used for Performance-Based Funding of Research Institutions in Norway", *European Political Science*, Vol. 8(3), pp.364-378.

Sivertsen, G. (2008), "Experiences with a Bibliometric Model for Performance Based Funding of Research Institutions", in Gorraiz J. and E. Schiebel (eds), *Excellence and Emergence: A New Challenge for the Combination of Quantitative and Qualitative Approaches,* Book of Abstracts, 10th International Conference on Science & Technology Indicators, Vienna, Austria.

Sivertsen, G. (2009), "A Bibliometric Funding Model Based on a National Research Information System", *www.issi2009.org/agendas/issiprogram/activity.php?lang=en&id=108*

Sivertsen, G. (2010), "A Performance Indicator based on Complete Data for the Scientific Publication Output at Research Institutions", *ISSI Newsletter*, Vol. 6(1), pp. 22-28.

Smart, W. (2009), *The Impact of the Performance-Based Research Fund on the Research Productivity of New Zealand Universities*, *www.educationcounts.govt.nz/publications /tertiary_education/49897*.

Strehl, F. (2007), "Funding Systems and Their Effects on Higher Education Systems – International Report", *www.oecd.org/publicationanddocuments/0,3395,en_33873108_33873681_1_1_1_9_ 1,00.html*

Tapper, T. and B. Salter (2003), "Interpreting the Process of Change in Higher Education: The Case of the Research Assessment Exercises", *Higher Education Quarterly*, Vol. 57(1), pp. 4-23.

UK Department of Health (2001), *Joint Funding Bodies' Review of Research Assessment: Response from the Department of Health to the Public Consultation Launched in September 2002*, *www.hefce.ac.uk/pubs/hefce/2001/01_64.htm*.

UK Government (2006), *Science and Innovation Investment Framework 2004-2014: Next Steps*, *www.dius.gov.uk/policies/science/science-funding/ten-year-framework/next-steps*.

University Grants Committee (UGC), Hong Kong, China (2006), *Research Assessment Exercise 2006 Guidance Notes, www.ugc.edu.hk/eng/ugc/publication/prog/rae/rae.htm*.

Whitley, R. (2008),"Changing Governance of the Public Sciences", in R. Whitley and J. Gläser (eds.), *The Changing Governance of the Sciences*, Sociology of the Sciences Yearbook, Springer, Dordrecht.

Whitley, R. and J. Gläser (eds.) (2008), *The Changing Governance of the Sciences*, Sociology of the Sciences Yearbook, Springer, Dordrecht.

Chapter 5

Highlights and reflections: Rapporteur's report

by

Marc Luwel
Centre for Science and Technology Studies (CWTS),
Leiden University, The Netherlands
and Hercules Foundation, Belgium

The OECD together with the Norwegian government organised the workshop, Performance-Based Funding for Public Research in Tertiary Education Institutions: Country Experiences, to do a stock-taking of the performance-based research funding systems (PRFSs) now in place in a number of countries. This chapter outlines the main topics discussed at the workshop, emerging key policy issues and possible next steps for work on this subject.

The changing role of universities

In industrialised countries, universities[1] have been profoundly modified over the last 50 years. They have moved from being rather elitist institutions to being key players in the knowledge society. They play three inter-connected roles: knowledge production, mainly through research; transfer of knowledge through education and formation; and the dissemination and application of knowledge through innovation. At the same time the emerging economies, especially the BRIC countries – Brazil, the Russian Federation, India and China – are rapidly expanding their higher education systems.

As the final stage of the formal educational system, the university still offers the highest level of education, but it is no longer for a small elite. The student population has increased rapidly, and its sociological and demographic composition has changed. Although research and scholarly work to advance knowledge and to fertilise teaching are still important aspects of the university's mission, applied research and outreach activities to strengthen national competitiveness and solve societal problems have taken on greater importance (Geuna, 1999).

Universities have evolved into multi-mission organisations operating in a complex and increasingly international context but still with a crucial local, regional and national role. As they become more entrepreneurial (Etzkowitz *et al.*, 2000), collaboration with industry and applications of research results have become additional sources of revenue to complement public funding from national, regional and supranational authorities and from charities.

The emerging knowledge economy has also had a profound impact on the public sector. Governments act less as operators and more as regulators and facilitators (Minc, 1985). As in many other sectors, public management reforms have afforded universities a certain degree of autonomy. At the same time, governments have introduced performance-based management and used market-type incentives to encourage universities to respond to governmental priorities. This has had a profound impact on the work environment of the academic staff, who often perceive these changes in terms of a loss of independence, a curtailment of academic freedom and a threat to the stability of research funding.

Performance-based research funding systems: a concept with many variations

While the United Kingdom's RAE is the oldest performance-based research funding system (PRFS), at least 13 countries have introduced such systems and selectively allocate institutional research funds to universities (see Chapter 1). Most countries explicitly or implicitly give a rationale for introducing a PRFS, the most frequent of which are to promote excellence through greater selectivity and concentration of resources and to better manage limited resources. The introduction of a PRFS is based on the assumption that it is possible to define research performance and, subsequently, to measure it. Performance is, however, a multidimensional phenomenon and is difficult to grasp.

Based on *ex post* evaluation, various aspects of research performance can be measured by indicators. These indicators can be classified in three main groups: first-order indicators directly aimed at measuring research performance by focusing on input, processes, structure and/or results; second-order indicators which summarise indexes in order to obtain simple measures for effect (*e.g.* journal impact factor and the H index); and third-order indicators from peer review panels that rate departments, for example. For quantitative indicators, data can be collected at any level; for practical reasons the peer review unit of analysis is the department or the field in the university.[2] The indicators are aggregated at university level for use in allocating block funding.

In most countries, the authorities have developed and often implemented the PRFS in close collaboration with the universities. However this did not always result in a large consensus on the indicators used in the different models. These indicators are in fact proxies that measure facets of a complex phenomenon. Critical comments, mostly formulated by academics, generally fall into two categories: the indicators themselves and their use in the funding formulas (see Chapter 4).

As research and innovation increasingly drove economies, science and innovation studies evolved into a mature research discipline, and sophisticated peer review methodologies and quantitative indicators were developed to evaluate and "measure" different aspects of the "business of science" and of science policy. However, it has become clear that there is no "ideal" methodology.

Peer review is the generic process of self-regulation of science and it provides indispensable credibility. Although it is held in high esteem by the academic community, it has limitations and potential biases (Cole *et al.*, 1981; Lawrence, 2003; Bornmann *et al.*, 2010). Quantitative indicators,

especially bibliometric indicators, become more robust at higher levels of aggregation; to evaluate individual (groups of) scientists they can only be used as background information for peer review. Moreover, the large, commercially available bibliographic databases such as the Web of Science or Scopus are not able to make scholarly work in social sciences and humanities and applied research sufficiently visible.

The intended and unintended consequences of PRFS are the subject of even more intense debate than indicators, although they are often intertwined. Distinctions must be made between reality and perception and between evidence-based and anecdotal evidence. As most systems were introduced at the end of the last and the beginning of this century, and taking into account that the impact of a PRFS is gradual, there is limited knowledge about their effects (see Chapter 2). Given its much longer history, it is not surprising that most studies are of the UK's RAE and that these provide the bulk of the available evidence.

As the foremost objective of the PRFS was to set up a funding allocation mechanism (partially) based on indicators of research performance, in order to make university funding (more) transparent and to make universities more accountable to the public authorities and the public at large, these objectives have been achieved. There are, especially in the United Kingdom, examples of management's response to or even anticipation of the introduction of a PRFS, such as departmental restructuring, strategic recruitment and a drive to create a culture of excellence. Another positive outcome in many countries is a significant improvement in the information management systems of universities or public administrations.

There is strong debate on the unintended consequences of PRFSs, with claims and counterclaims mostly based on anecdotal evidence. Quantitative and bibliometric indicators seem to generate the most aversion. A few often-heard criticisms and elements to refute them are:

- *The increase in scientific output is (largely) associated with game playing ("salami slicing" of publications).* However, no causality has been proven and the claim is counterintuitive as manuscripts go through a peer review process before they are published. Moreover, the combined use of publication and citation data in combination with journal impact factors would eliminate or limit possible biases.
- *Citations can be manipulated by citation fishing, citation cliques and self-citations.* The peer review process at journal level should be able to identify abusive use of self-citations and irrelevant citations.
- PRFSs have an impact on the disciplinary distribution of the research portfolio of research groups and of universities, owing to differences

in publication and citation culture between disciplines. The use of journal impact factors can correct for differences in citation culture. With Norway as the trail blazer, national or regional bibliographic databases have been set up to better cover scholarly work in humanities and social sciences.

- Excessive emphasis on publications and citations stimulates risk averseness, and hinders blue-sky, interdisciplinary and collaborative research, and training of PhD students. Bibliometric studies show that (international) collaborative research and interdisciplinary research have greater visibility than research carried out at a single institution or disciplinary work. Risk averseness would reduce the scientific relevance of results. Even if the work is accepted for publication, it would receive fewer citations. To reward the investments made in training PhD students, some governments use numbers of PhD students or of PhD degrees awarded as indicators.

- *The use of publication and citation data hampers collaboration with industry and other outreach activities.* Scientifically outstanding groups often also collaborate closely with industry. Moreover, governments can use indicators for industrial collaboration and application of research results.

For PRFS based on quantitative indicators, data on individual researchers' work are collected, but these data are amalgamated at institutional level and used to allocate fractions of the lump sum among universities based on relative performance. The management allocates the lump sum internally based on the university's mission statement and its priorities but within the regulatory framework laid down by the government. Some criticisms are often an appeal for stronger institutional management to counterbalance perceived or real negative consequences of PRFS.

As experience was gained with PRFS, the authorities often took criticisms into account, without always seeking proof of the allegations. Adjustments were made to correct for real or alleged biases. The UK Higher Education Funding Council of England (HEFCE), for example, modified the assessment methodology based on the evaluation of successive RAEs. In some cases, additional indicators were introduced to reflect new governmental priorities. These modifications and additions often increased the complexity of the system and the overall cost of managing it, in some cases to the detriment of consistency.

Some policy issues

Over the next few years universities will be confronted with challenges on the content of their "social contract". These challenges are already apparent in discussions about the balance between education, research and outreach activities, between private and public funding, between blue-sky research and commercially oriented research, and between institutional autonomy and accountability (Zusman, 2005).

Given the impact of the present economic downturn on public expenditures and the societal implications of new challenges such as an ageing population and the effects of globalisation on economic competiveness, university funding will be under pressure. Universities will be asked to do more with less and to demonstrate that they use the available public funds effectively and efficiently. Universities become increasingly "privatised" as they must look for additional funding outside the public sphere, especially through industry funding for academic research and university/industry partnerships. Although governments remain an important funding source, these developments have implications for the relationship between universities, the public authorities and the public at large as well as for academic staff (Geiger, 2004).

But public management of the higher education sector and the cost of managing funding models will also be scrutinised as governments try to reduce their administrative costs.

A possible research agenda to develop more robust PRFS

The literature contains few studies that examine in detail the impact of PRFSs by using data analysis or well-structured surveys and qualitative investigations. Such studies must demonstrate causality between changes in funding and systemic effects visible at the national, institutional, departmental and individual levels. Taking into account all competitive (and non-competitive) funding sources, the regulatory framework for the higher education sector, all the components of the universities' mission and the relatively small fraction of the lump sum allocated using PRFS, this is a daunting task (Lane, 2010).

To mention only a few topics to be investigated:

- The interaction of PRFS with the educational mission and outreach activities of universities.
- The relationship – if any – between the size of the lump sum (re-) allocated based on PRFS and its impact at the national, regional, institutional and departmental level.
- The relationship between the total cost to set up and manage the PRFS and its economic and societal benefits.

- The effects of PRFS on the university management as at least a fraction of the funding is available for discretionary spending, taking into account the regulatory framework in which it has to operate.
- If selectivity and concentration are objectives, is there an optimum to be realised with PRFS?
- Is the mere existence of PRFS a strong enough incentive to stimulate performance or is there an optimal fraction of institutional and/or total public research funding to be allocated using PRFS?
- Is the use of PRFS creating a more or a less attractive research environment for young researchers and tenured staff (brain drain, brain gain)?
- Is the allocation of funds based on relative rather than absolute performance the most appropriate? PRFS based on relative performance can counterproductive as they may not reward improved potential.
- Given the fast development of data mining and full text search tools, is it possible to develop and to introduce more sophisticated measures, such as citations to sub-sections of papers (Campbell, 2008), download statistics of documents, and contextual citation impact of journals (Moed, 2010) and to design better information management tools?[3]
- Especially for smaller countries, is benchmarking universities within the national system the most appropriate method?

It is true that a good metric is difficult to develop, but it is not a reason not to take up the challenge in order to make better funding decisions, stimulate excellence and reward the best scientists. It is preferable to carry out this work in a transnational setting and in close collaboration with universities. As indicated above, a group of OECD countries have set up PRFS with fairly similar objectives. Each system is embedded in its national context and is implemented differently. International comparative studies on PRFS demand in-depth knowledge of the national systems. Much of the indispensable information is not available in the scholarly literature and is buried in legal documents and grey literature written in the national languages. Benchmarking should not be limited to countries that have implemented PRFS and should also include countries with a large number of private universities.

The studies' outcomes could be of great benefit to national authorities and universities in their efforts to increase the effectiveness and the efficiency of institutional funding, not only for countries using PRFSs but also for countries interested in setting up such systems.

Notes

1. Although private higher education institutions play an important role in some countries, the term "universities" is used here to designate public higher education institutions and publicly funded higher education institutes.
2. However, the Spanish *sexenio* and the New Zealand PBRF grade individual researchers and their research record. The sexenio is used to set the researcher's salary.
3. See, for example, *http://lattes.cnpq.br/english/index.htm*.

References

Bornmann, L., L. Leydesdorff, and P. Van den Besselaar (2010), "A Meta-Evaluation of Scientific Research Proposals: Different Ways of Comparing Rejected to Awarded Applications", *Journal of Informetrics*, Vol. 4, pp. 211-220.

Campbell, P. (2008), "Escape from the Impact Factor", *Inter-Research*, Vol. 8, pp. 5-7.

Cole, S., J.R. Cole and G.A. Simon (1981), "Chance and Consensus in Peer Review", *Science*, Vol. 214. pp. 881-886.

Etzkowitz, H. *et al.* (2000), "The Future of the University and the University of the Future: Evolution of the Ivory Tower to Entrepreneurial Paradigm", *Research Policy*, Vol. 29, pp. 313-330.

Geiger, R. (2004), *Knowledge & Money – Research Universities and the Paradox of the Marketplace*, Stanford University Press, Stanford, CA.

Geuna, A. (1999), *The Economics of Knowledge Production*, Edward Elgar, Cheltenham.

Lane, J. (2010), "Let's Make Science Metrics More Scientific", *Nature*, Vol. 464, pp. 488-489.

Lawrence, P. (2003), "The Politics of Publication", *Nature*, Vol. 422, pp. 259-261.

Minc, A. (1985), *L'Etat Minimum*, Albin Michel, Paris.

Moed, F. (2010), "Measuring Contextual Citation Impact of Scientific Journals", *Journal of Infometrics*, Vol. 4, pp 265-277.

Zusman A. (2005), "Challenges Facing Higher Education in the Twenty-First Century", in P. Altbach, R. Berdahl and P. Gumport (eds.), *American Higher Education in the Twenty-First Century. Social, Political, and Economical Challenges*, The Johns Hopkins University Press, Baltimore, MD.

Annex A

Workshop summary

by
Sarah Box
OECD Secretariat

The workshop on performance-based funding of public research in tertiary education institutions was opened by **Mr. Andrew Wyckoff** (Director, OECD Directorate for Science, Technology and Industry) and **Ms. Toril Johansson** (Director General, Norwegian Ministry of Education and Research). Mr. Wyckoff welcomed participants and highlighted the crucial contribution that the research of tertiary education institutions (TEIs) can make to solving some of society's big challenges. Ms. Johansson noted that ongoing changes to countries' funding schemes suggested that the ideal model had not yet been found, and expressed hope that the workshop would help in gaining more knowledge about systems and performance indicators as well as in identifying important issues for further exploration. The meeting was then handed over to the day's chairwomen, **Ms. Karen Haegemans** (Flemish Authority, *Department Economie, Wetenschap en Innovatie*, Belgium) and **Ms. Jana Weidemann** (Ministry of Education and Research, Norway).

The workshop began with a keynote address by **Mr. Aldo Geuna** (University of Torino, Italy). Mr. Geuna set the scene by highlighting how the massification of higher education, resulting in larger organisations and many new actors, has generated new approaches to governance and accountability, including performance-based funding. Assessing these schemes is difficult, as they generally form just one part of a bigger funding profile that includes many funding sources (both competitive and non-competitive). Citation counts are one performance indicator that may be used within performance-based funding schemes, and it is a fast-developing field. However, Mr. Geuna questioned whether citations can accurately capture differences in quality at different points in the distribution (*e.g.* what is the difference between 350 and 550 citations, and is the marginal difference consistent over the spectrum?). He also questioned whether top scientists required incentives to work "more". Nevertheless, citations may be useful as an indicator of low research performance, and could be used as a "hurdle" for being considered as an active researcher. Mr. Geuna noted the pressures on salaries and the price of services provided by academics that can appear within performance-based funding models, as more demands are placed on their activities.

Ms. Diana Hicks (Georgia Institute of Technology, United States) then presented her paper, "Models of Performance-Based Research Funding Systems" (see Chapter 1). Ms. Hicks highlighted the various rationales for such systems, including rewarding manifest performance, contributing to economic performance via quality and excellence in research, implementing public management reforms such as increasing accountability, and meeting country-specific goals such as an improved international research profile. After presenting a classification of performance-based funding systems, Ms. Hicks noted the uncertain but potentially high costs of some systems, and the variety of different estimates put forward for their financial impact. She also

pointed to some dynamic considerations, such as the increasing complexity of some models and the question of whether there are diminishing returns. Ms. Hicks concluded that performance-based research funding systems might be good for introducing contestability to encourage research excellence, but may not be so good for fostering equity, maintaining national and cultural identity, encouraging broader societal and economic outcomes for research, or supporting novelty and intellectual diversity. She called for comparative research into alternatives to performance-based funding, such as international university rankings and centre of excellence approaches.

The paper by Ms. Hicks was discussed by **Mr. Dominic Orr** (*HIS Hochschul-Informations-System GmbH*, Germany), who agreed that both the higher education and innovation studies literature streams are important for the analysis of performance-based funding schemes. He also highlighted the importance of international studies that can stimulate cross-country debate and comparisons. Mr. Orr commented that inherent conflicts within performance-based procedures contribute to the high level of diversity between countries and to the instability of models within countries over time. Such conflicts include tensions between the complexity and practicality of models, between autonomy of institutions and control from outside, between institutional expectations of reward for performance and the realities of government budget constraints, and between the priorities set by the performance model and those set by other funding tools. Mr. Orr thought it unlikely that existing information systems such as university rankings could fully substitute for a performance-based scheme, as the priorities of the authorities responsible for national higher education are unlikely to be equivalent to those highlighted in ranking mechanisms.

Ms. Hanne Foss Hansen (University of Copenhagen, Denmark) then presented her paper, "Performance Indicators used in Performance-Based Research Funding Systems" (see Chapter 2). Ms. Hansen described the rich world of indicators and its ongoing enlargement resulting from the search for improvements to the quality and information content of performance indicators. She classified indicators into three groups: first-order indicators, related to inputs, processes, structures, outputs and effects; second-order indicators, which tend to be summary indexes; and third-order indicators, which revolve around peer review. Country systems today present examples of both first-order and third-order indicator models, as well as mixed models. Ms. Hansen noted that good indicators should meet a number of criteria, including being clear and understandable, relevant, verifiable, credible and available at reasonable cost. However, there is little systematic knowledge of how indicators really "shape up" against some of these criteria. Ms. Hansen concluded by calling for more knowledge on how contexts and actor strategies shape

performance-based funding schemes and how these then feed back to actors' strategies and behaviours.

Ms. Hansen's paper was discussed by **Mr. Armel de la Bourdonnaye** (*Ministère de l'Enseignement Supérieur et de la Recherche*, France), who drew on France's recent experiences with peer review mechanisms to raise some issues and questions. Mr. de la Bourdonnaye noted that performance-based systems are based on a set of beliefs that merit further discussion. One belief is that institutions will try to maximise their funding and that their performance in the future will be a prolongation of their performance in the past. Another belief is that performance-based systems optimise the allocation of resources, supposing that the marginal gains from increased funding are greatest in those institutions that have the highest measured performance. A third belief is that a multi-criteria analysis (combining, say, research strategy and research excellence) can be transformed into a single indicator (money) without difficulty. Mr. de la Bourdonnaye posed a number of open questions, such as how to implement national strategic orientations via a performance-based scheme, how to fund risky initiatives, and how to avoid "strategic behaviours" that lead research institutions to abandon non-measured missions (such as public outreach). He also questioned the size of the global added value of performance systems if there is not globally more money available. Finally, Mr. de la Bourdonnaye asked how in countries such as France, where some institutions are submitted to a performance scheme and others are not, a dual system can be managed efficiently.

The next presentation was of the results of the RIHR country survey (see Chapter 3), presented by **Ms. Sarah Box** (OECD Secretariat). Ms. Box noted that institutional funding streams are a large proportion of total funding in many countries, so that improving the efficiency and effectiveness of this spending is important. The survey responses revealed that most current schemes for performance-based funding of public research in TEIs have been introduced since 2000. Stimulating quality research is a central rationale, but countries also seek to achieve other goals with their performance schemes. Ms. Box found that most schemes used similar indicators of performance, although in quite different combinations and with different weights, and that the budgetary implications of assessments varied widely across countries. Few of the schemes had been formally evaluated, but existing evidence suggested that the schemes had positive effects on research output and research management, although "cause and effect" was difficult to establish. Ms. Box concluded by suggesting some possible issues for discussion, including whether multiple goals for performance-based funding schemes are workable and appropriate, which indicators have the greatest impact on incentives, and how co-ordination with other funding streams could be improved.

Ms. Linda Butler (Australian National University and University of Newcastle, Australia) then presented her paper, "Impacts of Performance-based Research Funding Systems" (see Chapter 4). Ms. Butler highlighted the difficulties of assessing the impact of these systems, particularly in distinguishing between intended and unintended consequences and between reality and perception, and in establishing whether outcomes are desirable or not. She also noted that demonstrating causality is hard, and presenting an evidence-based assessment rather than anecdotes is a challenge in this area. Ms. Butler found that most countries had achieved their intended outcomes, in terms of distributing funding, improving research quality and increasing accountability for government funding. She then discussed the available evidence on some unintended effects that are the source of concerns in academic and policy circles, such as choice of publication outlet, staff morale, teaching, discipline mix, types of research, collaboration and institutional management practices. Ms. Butler noted that most of the evidence was anecdotal or survey-based, although she pointed out that bibliometric analysis could provide insight into some of the issues. Ms. Butler questioned whether impacts are due to specific systems, or whether they result from the simple existence of a system, and whether some behavioural responses should be seen as positive or negative. She called for relevant unpublished analyses (such as internal government documents) to be made public, in order to provide more evidence about performance-based funding systems, and for multi-national multi-dimensional studies to be set up.

Ms. Butler's paper was discussed by **Mr. Gunnar Sivertsen** (NIFU STEP, Norway). Mr. Sivertsen commented that "now we know how little we know", and agreed with Ms. Butler's suggestion to have internationally co-ordinated studies that make use of comparative perspectives and data. Mr. Sivertsen raised three questions that arose from the material collected and presented by Ms. Butler. The first concerned the roles of researchers doing research on their own research conditions. Mr. Sivertsen noted that researchers may find themselves playing the roles of "victims", "witnesses" or "investigators", and that it is important to distinguish between reality and perceptions, and opinions and fact. He called for new research based on best practice social science methods, performed independently and with a clear understanding of roles. Mr. Sivertsen's second question concerned the perspectives that studies present on change, noting that change often seems to be regarded as negative. He suggested that change should be expected, and that origins, contexts, processes and results of change should be studied with an open mind. Finally, Mr. Sivertsen questioned why so many studies focused on the individual researcher rather than the institution, even though funding models are aimed at institutions. He contrasted this with higher education studies which tend to favour an organisational perspective, and called for a shift in perspective in

future studies to highlight the institutional changes that take place under performance-based funding.

Following the presentation of the four project work blocks, the workshop moved to a panel debate, led by **Ms. Rosa Fernández** (Department for Business Innovation and Skills, United Kingdom). Panel members were **Mr. Bjorn Haugstad** (University of Oslo, Norway), **Mr. Paul Hubbard** (Higher Education Funding Council of England – HEFCE, United Kingdom), **Ms. Julita Jablecka** (University of Warsaw, Poland) and **Mr. Poul Erik Mouritzen** (University of Southern Denmark, Denmark). The panel members were asked to discuss the statement: "Performance-based funding models are the most efficient and effective way for governments to drive their public research goals in tertiary education institutions".

In his opening comments, Mr. Haugstad claimed that while the statement was a bold assertion, performance-based funding may be the "least bad" option, in recollection of Winston Churchill's statement that "democracy is the worst form of government except all those other forms that have been tried from time to time". [1] The issue, then, is not whether to have a performance-based research funding system, but what it should look like. In Norway, performance-based funding has increased productivity (in terms of number of articles), has redistributed money to success, has increased attention to publishing and publication strategies in academic departments, and has raised awareness of who publishes and who does not. However, Mr. Haugstad questioned whether the main impact may be to raise standards from medium to medium-high, rather than to excellent. He also suggested that institutions may have to constantly raise their performance simply to stay at the same funding level (a "red queen" game), and that the capacity of institutions to change in response to performance assessments may be limited due to government regulations and rigid political stakeholders. Mr. Haugstad concluded by questioning the interplay between institutional and project funds and the impact this has on institutions.

Mr. Hubbard claimed that the statement was true, up to a point. He suggested that performance-based funding systems are a strong tool for delivering national policy aims when driving research excellence is an essential requirement and limited funds have to be allocated to best effect. However, he suggested they are less well suited for steering the subject mix of the research base or encouraging an optimal distribution of PhD students. In Mr. Hubbard's view, successful systems are highly visible, transparent, well understood and well respected, and are done "by" researchers more than "to" them. He noted that systems must be carefully designed to fit policy priorities and must be flexible enough to recognise and reward a broad range of research

1. Hansard (United Kingdom House of Commons), 11 November 1947.

types and approaches (including variations among disciplines). Performance-based research funding systems can strengthen institutional autonomy, promote a culture of excellence and efficiency, and stimulate faster and more effective dissemination of outputs. However, they can also be costly and complex, and be used as a "scapegoat" for any general problems within the research system. From the United Kingdom's experience, quality and productivity will result over time, although some evolution in the system will undoubtedly be necessary. Reasonably small marginal shifts in funding may be enough if researchers believe that assessment outcomes matter.

Mr. Mouritzen believed the statement could be answered by a "no", an "it depends", a "yes" or a "maybe". For some organisations, the introduction of performance-based models will have few, if any, positive effects. This is particularly the case for institutions in which motivation is based on intrinsic rewards, promotion is already based on transparent quality criteria, researchers already work hard and quality is in the top echelons. In other institutions, the perceptions of staff will be determinant – if the system is perceived as a management control tool, then performance will drop, but if it is perceived as a supportive measure, performance may improve. To support a "yes" response, Mr. Mouritzen commented that when organisations are exposed to evaluative indicators, researchers will respond and will seek to increase the value of the indicator. Finally, Mr. Mouritzen noted that researchers implement "coping strategies" to deal with performance systems and, depending on how these work, the publication strategy and thus performance of the institution will be influenced in different ways. The fact that the statement can be responded to in a number of ways suggests that there is still much to learn about performance-based funding systems, and Mr. Mouritzen called for more systematic knowledge.

Ms. Jablecka agreed that performance-based funding systems are efficient and effective, but stressed that goals must be clear. An overly complex system is not transparent and the relationships between indicators and induced performance become difficult to identify. Ms. Jablecka commented that performance systems do not always lead to increased quality, for example when the benchmark is national rather than international performance. She questioned whether systems that count publications are able to enhance creativity, and suggested that stable long-term financing is better suited for enabling big discoveries. Ms. Jablecka suggested that performance systems may be biased against the restructuring of research units, as promising young researchers who do not yet have a track record are less likely to be hired. Finally, Ms. Jablecka noted that autonomy is a very country-specific notion and perceptions depend on the arrangements that institutions have been subject to in the past.

The panel's comments spurred a discussion of the quality of research work. Participants questioned whether an increase in citations indicates that quality has risen, and noted that it is important to look at exactly where citations and other performance indicators have risen (*i.e.* in the tail of the distribution or in the top 1-5% of articles?) and not just at averages. It was noted that changes in countries' citation impacts did not seem to be associated with changes in performance systems (although behaviour may change in response to anticipated system changes) and that country contexts were extremely important. The "salami slicing" phenomenon (where research is parcelled into the smallest possible publishable pieces to maximise the number of articles), if it exists, was considered to reflect inadequacies in peer review processes. Participants suggested that it is still not particularly acceptable for researchers to act in this way, although attitudes may change in response to performance incentives. It was also suggested that creating institutions that nurture quality is of great importance, and this requires good leaders who provide motivation and incentives.

Other questions raised by participants included when the return to performance assessment becomes flat, that is, at what point does it bring no further benefits? It was also noted that any funding system generates incentives and it is important to consider what these are in evaluating different options. Participants discussed whether risky and ground-breaking research would be better supported by other funding options, such as peer review or centres of excellence, although this was not seen as a reason not to use performance systems. Having a mix of funding streams that could be aimed at goals such as support of young researchers or interdisciplinary research could complement performance schemes.

To draw together the day's debate, the workshop rapporteur **Mr. Marc Luwel** (Leiden University, the Netherlands and Hercules Foundation, Belgium) presented a short review of his thoughts on the main issues and policy debates around performance-based funding for public research in TEIs (see Chapter 5). He noted the evolution of universities over the past few decades, towards more autonomy, but with accompanying checks. Universities operate in a complex funding environment, with different levels and sources. They are also embedded in national and regulatory contexts. In the current environment, there are competing priorities for taxpayer dollars, and there is the chance that higher education funding will come under increasing pressure.

Mr. Luwel noted that performance systems have evolved in a step by step fashion, which has led to some complexity, with questionable results for efficiency. Given that funding is for institutions, there needs to be better analysis of the impact on institutions. Improved policies within organisations can be crucial for performance, although Mr. Luwel questioned whether strong institutional leadership is sufficient to avoid potential negative consequences of

performance systems. On funding, Mr. Luwel asked whether a fixed pool of funding that is shared among institutions is the best approach, and whether all improvements should be rewarded (rather than simply the largest). In conclusion, Mr. Luwel commented that since performance-based systems are probably here to stay, and will require assessment over time, there is a need for multinational assessments, undertaken by international experts with in-depth knowledge, to help benchmark different systems.

Main discussion themes

The open discussions that took place over the course of the workshop were wide-ranging. One interesting debate concerned the interaction of funding streams, in particular whether they complement one another or whether they have contradictory effects. Participants tended to agree that it was not possible to achieve all goals via one funding stream, and that the allocation of funding must also recognise the dual role of universities in research and teaching. A combination of strategic funding (giving a forward-looking perspective), performance-based funding (looking backwards), and a portion of funding based on previous years' allocations (for stability) could be one approach, for example. This raises the importance of understanding the relationships between funding streams and co-ordinating them so that incentives are reasonably well aligned. Delegates noted that the relationship between government funds and other funding sources was a crucial part of this. At a general level, it was considered vital to be very clear about the types of performance governments are seeking and exactly what they wish to fund (*e.g.* research, institutions, etc.), so as to identify the best mechanism to achieve public policy goals. Understanding what funders and academics consider to be "performance" is important for designing good funding mechanisms.

Related to this, delegates noted that within performance-based systems, different indicators and measures of performance give different information. This should be recognised and they should be used in complementary ways. At the same time, an eye should also be kept on the costs of funding systems. Some delegates questioned whether individual research performance should be at the base of models, since researchers are the building blocks of the research system. It was noted that models take a mixed approach to the level of assessment, and in those countries where individual performance is not targeted by the performance-based funding system there may be other mechanisms (*e.g.* research council funds) that systematically review individual performance. There was a question as to whether the effects of "rewarding" or "punishing" performance would be symmetrical under a performance-based scheme; it was noted that systems essentially do both, by reducing funding to the lower tail and increasing funding at the top.

Another area of discussion was the required size of monetary incentives for schemes to have an impact. It was noted that even implementing performance-based funding on a 2% slice of institutional funding can have an effect, although in some countries, reputation and peer pressure creates sufficient incentives for good research. It is necessary to distinguish between the total amount of funding at stake and the marginal funds that move as a result of assessment, with small marginal shifts able to have powerful effects. At the same time, the allocation mechanism may be as important as the amount of funding (*e.g.* "winner takes all" versus proportional allocations). The way in which funds are used within universities was also raised as an important factor in how performance-based funding affects research. Incentives for risky research were also discussed, and it was noted that there was no "utopia" of innovative risk-taking prior to the implementation of performance-based funding schemes and that peer review does not necessarily do a better job of supporting such research.

Quality was discussed, with some participants noting that nationally valuable research may not necessarily be regarded as useful at an international level. For this reason, systems of international ranking would not provide the appropriate incentives for institutions, and nor would centre of excellence funding streams if the research milieu is small. Performance-based funding can thus help to promote quality in small fields. The debate about quality also highlighted the importance of having a mixture of funding streams.

In terms of possible future analytical work, delegates agreed that international comparative studies would be valuable, particularly those that could identify general developments affecting all countries as well as specific country trends.





OECD-Norway workshop on performance-based funding for public research in tertiary education institutions

Project under the OECD Working Party on Research Institutions and Human Resources (RIHR)

21 June 2010

WORKSHOP AGENDA

MONDAY, 21 JUNE 2010	
09:30-09.45	**1. Welcome** ***Mr. Andrew Wyckoff*** (Director, Directorate for Science, Technology and Industry, OECD); and ***Ms. Toril Johansson*** (Director General, Norwegian Ministry of Education and Research) *Introduction to the day's proceedings by the OECD Secretariat.*
09.45-10.15	**2. Keynote address** ***Mr. Aldo Geuna*** (University of Torino, Italy)
10.15-11.15	**3. Models of performance-based funding** Presentation by ***Ms. Diana Hicks*** (Georgia Institute of Technology, United States) *Discussant:* ***Mr. Dominic Orr*** (HIS Hochschul-Informations-System GmbH, Germany)
11.45-12.45	**4. Indicators for performance-based funding** Presentation by ***Ms. Hanne Foss Hansen*** (University of Copenhagen, Denmark) *Discussant:* ***Mr. Armel de la Bourdonnaye*** (Ministère de l'Enseignement Supérieur et de la Recherche, France)
14.15-15.00	**5. Country survey results** Presentation by the ***OECD Secretariat***
15.00-16.00	**6. Impacts of performance-based funding** Presentation by ***Ms. Linda Butler*** (Australian National University, Australia) *Discussant:* ***Mr. Gunnar Sivertsen*** (NIFUSTEP, Norway)

16.20-17.20	**7. Panel debate/roundtable** *Panel debate topic:* "Performance-based funding models are the most efficient and effective way for governments to drive their public research goals in tertiary education institutions." *Facilitator:* ***Ms. Rosa Fernández*** (Department for Business Innovation and Skills, United Kingdom) *Panel members:* ***Mr. Bjorn Haugstad*** (University of Oslo, Norway) ***Mr. Paul Hubbard*** (Higher Education Funding Council of England – HEFCE, United Kingdom) ***Mr. Poul Erik Mouritzen*** (Department of Political Science and Public Management, University of Southern Denmark, Denmark) ***Ms. Julita Jablecka*** (Centre for Science Policy and Higher Education, University of Warsaw, Poland)
17.20-17.35	**8. Rapporteur's summary** ***Mr. Marc Luwel*** (Herculesstichting, Belgium)
17.35-17.45	**Close of workshop**

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where governments work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Commission takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

OECD PUBLISHING, 2, rue André-Pascal, 75775 PARIS CEDEX 16
(92 2010 02 1 P) ISBN 978-92-64-09460-4 – No. 57219 2010

Performance-based Funding for Public Research in Tertiary Education Institutions

WORKSHOP PROCEEDINGS

Governments are seeking to channel research funds into universities and other institutions in ways that support high-quality research in strategically important areas and bolster effective knowledge diffusion. These issues of steering and funding have even more relevance in light of the current financial crisis and economic downturn which have seen severe fiscal pressures fall on many countries.

This publication presents a collection of papers presented at an OECD workshop on performance-based funding for public research in tertiary education institutions. It takes stock of current thinking and practice around performance-based funding for public research in tertiary education institutions, as a tool to help governments meet their research goals. These funding models are essentially systems of *ex post* evaluation of research outputs and outcomes from universities and other tertiary institutions, and are generally based on peer review, bibliometrics or other quantitative indicators. Their results are used to inform government decisions about how much and which institutions to fund.

Please cite this publication as:

OECD (2010), *Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings*, OECD Publishing.
http://dx.doi.org/10.1787/9789264094611-en

This work is published on the *OECD iLibrary*, which gathers all OECD books, periodicals and statistical databases. Visit ***www.oecd-ilibrary.org***, and do not hesitate to contact us for more information.

OECD*publishing*
www.oecd.org/publishing

ISBN 978-92-64-09460-4
92 2010 02 1 P

