OECD ISSUE BRIEF: RESEARCH ORGANISATION EVALUATION

The issue in a nutshell

This brief focuses on the evaluation of research organisations. Evaluation is a policy tool which is used to steer, manage and improve the activities of and investments in public sector research organisations. It is also used to change the distribution of funding among research organisations. Modes and mechanisms of evaluation have been developing over recent decades as budget holders have demanded accountability for public money spent, in common with other areas of public expenditure. In addition to simply justifying research budgets at a political level is the desire for evaluation to reveal the breadth of impacts of research, and the relative academic quality of research organisations and the research groups within them, with reference to both the national and the international frame. It sits alongside performance measurement systems, monitoring and performance indicators which are commonly implemented for the management and control of public sector bodies.

Evaluation of research holds particular challenges. First is the need to involve experts in order to judge the quality of research. This brings in the need for peer review. Second is the unpredictable nature of research outcomes and their application, including the fact that only a few lines of research will ultimately have any significant impact. The length of time required to see impacts and effects can be beyond that which is useful for present-day management and policy decisions. Impacts are usually not directly attributable to one piece of research. As a result, traditional economic techniques under-value the impact of research. Yet public research organisations are increasingly aware that they must demonstrate performance, impact and quality to their parent funding bodies, to their private clients and to the international research community.

We see in several national contexts the implementation of national research evaluation systems which scrutinise research organisations. These have also been called “performance based research funding systems”. Whitley has characterised these as “strong” and “weak” according to the consequences of evaluation in terms of resource allocation. Strong evaluation systems redistribute resources according to the evaluation results. Here the policy discourse is about concentrating “scarce” funding in the “best” groups and institutions and removing it from “poorly performing groups. Weak evaluation systems have little or no impact on resource allocation but give reputational signals. We can see that “weak” evaluation systems nevertheless mobilise public research organisations to do as well as possible, because reputation is a key currency for researchers and their institutions. Incentives for organisations to take part in evaluation need not be about financial rewards alone.

In other contexts we see the adoption of evaluation alongside strategic planning at the level of individual organisations. Peer/expert review-based evaluation is used to aid institutional renewal or re-positioning. This is often combined with a focus on the performance and career progression of individual researchers. Public research organisations may use essentially a programme evaluation framework (time bound and incorporating targets and objectives) to examine the lines of research which their groups are pursuing. They assess their relative success in an internal competition for funding and group survival which in turn shapes the overall institution and its strategic planning until the next evaluation.

At another end of the spectrum, alongside linking evaluation to organisational strategy and planning processes, there has been an increasing interest in linking to system level evaluations of research and innovation policies. The view here is that public research organisations are part of a complex web of institutions creating and using knowledge and that the ultimate innovative performance of the system is dependent upon the interactions and all the myriad policies which affect them. A classic tension can be seen when we look across national contexts: evaluation to assure public accountability at a high level of aggregation, for example block grants to institutions versus the development of evaluation processes which foster learning and organisational development and improvement of the quality and impacts of research.

The international frame is becoming increasingly important as research itself is ever more international (cooperative programmes, mobility of researchers and the increase in internationally co-authored papers). Evaluations of public research organisations often provide data for benchmarking of national performance. This led to the proliferation of league tables, for example universities and research institutions and groups within them, mainly according to bibliographic data on publications and citations. The research organisations and their national or regional funding bodies are increasingly aware of the reputational importance of their positions in league tables.

The degree to which research organisation evaluation influences innovation depends on the criteria for evaluation. Where they stress international research excellence as judged by peers, the incentive is towards academic publishing and away from applied research and links with industry. Where the ability of research organisations to attract private sector funding, place patents and foster spin-off companies is counted and assessed, there will be incentives for this kind of engagement. Public sector research systems (including the different kinds and levels of evaluation of public research organisations) may include incentives for both highly academic research and industry/societal relevance.
Who are the main actors in research evaluation?

It is useful to separate the actors involved into those commissioning or initiating evaluation, those being subjected to evaluation, actors performing evaluation and audiences for evaluation. The main actors in the public research system are relevant: research performers, research funders and intermediary organisations delegated to allocate research funding. The picture is complex. Evaluations may be initiated by overall budget holders, that is governments – federal, national and regional - private foundations and charities. These actors will typically have an interest in demonstrating that research budgets have been properly spent and show value for money and impact. Evaluation will probably sit alongside routine audit and performance/financial reporting processes. These funding bodies have the power to initiate new forms of evaluation and similar activities such as benchmarking exercises. An example here would be the initiation of national research evaluation exercises such as the research assessment exercise in the UK and similar ones in Norway, Australia and the Netherlands. Research organisations themselves may initiate evaluation, perhaps to pre-empt an undesirable form of evaluation imposed from above, or as part of strategic re-alignment, justifying their existence and providing evidence of their quality and worth.

Subjected to evaluation are the performers of research: universities (Higher Education Institutions), national public research organisations, government laboratories, which usually perform mission-oriented research and international organisations such as the European Space Agency and the research centres of the European Union. The composition of research performers varies considerably between national settings for example, universities being the prime sector for undertaking research versus national public research organisations or government laboratories. Organisations subjected to evaluation develop responses and behaviours to mitigate evaluations, such as forming new internal processes for review and data collection in order to promote the best possible evaluation results.

Intermediary organisations which allocate research funding on behalf of a Government Department play an important role in evaluation but have, with rare exceptions, not been subject to evaluation. They may operate evaluation of their grants and programmes and, if applicable, of their own Institutes. We find specialist evaluation Agencies and Commissions in settings where performance based funding exists. Actors in the public research system have been diversifying with typically more complex relationships between them and thus one type of organisation may have multiple roles in evaluation, for example, initiating a self evaluation and being subjected to an external evaluation. Finally, individuals are important where evaluation at individual level overlaps with organisation evaluation, and in individual adaptations to performance regimes.

---

Actors involved in performing evaluation are organisations and individuals. As mentioned above, there exist intermediary public agencies for executing evaluations and collecting statistics including bibliometric data. They will almost certainly be operating peer and/or expert panels to generate evidence for evaluation. Many research organisation evaluations incorporate a large element of self-reporting, that is compiling evidence and self-assessments for peers to judge and in this sense research performers usually also evaluation performers, although not making the final judgements about themselves. Their individual members will probably be asked to act as peer reviewers in the evaluation of other organisations. In the related area of programme evaluation, there exist numerous private consulting firms, both the well-established large firms and small specialist firms. These may be hired to provide specific evidence as part of evaluation or indeed to execute organisational evaluations in full.

The audiences for evaluation include, in theory, the general public (voters), politicians and policy-makers at large. In practice, evaluation of research organisations is of most interest to the main actors in the research system, Ministries or departments for research, education and industry, and industry itself.

**What are the main mechanisms?**

Before considering examples of mechanisms, it is worth briefly understanding the salient features of peer review and STI indicators, since most research organisation evaluations use at least one of them. Both require institutional arrangements to produce evaluation evidence that is credible and robust.

**Peer Review**

Peer review refers to the judgement made on the quality of research (or individual researcher) by someone who is knowledgeable and expert in the area. The peer has to be familiar with the state of the field, the main actors within it, the frontiers of the field and what researchers within it find interesting. They make judgements about quality which incorporate notions of robustness of knowledge, novelty and originality and likelihood of success of research. They may also be asked to judge the relevance and applicability (also called extended peer review). Peer review brings practical problems of finding disinterested peers, which can be difficult in a small field or small country. It also relies on the good will and input of time of peers, particularly if site visits and interviews with research organisations are to take place. Peer review has some well-known limitations: it tends to favour high quality but conservative research rather than break-through research and novelty. It tends to favour established groups rather than new researchers. Funding agencies may worry that it is too “friendly”. Peer review needs careful organisation to ensure that criteria for judgement are clear and consistent, for example by defining scores. Finally, employing expert researchers in peer review is very costly not only in direct costs of travel and meetings, but in indirect costs – diverting senior people from their main functions.
Indicators

The use of publication and citation counts has become widespread due to the availability of databases and techniques for collecting and analysing data. Indicators are used at country level, for example to look at relative share of world publications and citations. They are increasingly being used in research organisation evaluation. They are seen as “neutral” compared to a peer judgement, can generate comparative performance data and look at first glance to be a cheaper way to measure quality than direct peer review. In fact, publication data can be seen as “indirect” peer review, since journal articles are accepted only after peer review. So publication indicators carry all the weaknesses of peer review.

Citations are mistakenly correlated with “quality”. We know very little about why articles are cited. It is more accurate to speak of citation “impact”, but still inexact. We do not know whether an article with 100 citations is “better” than one with 50 citations. Citations are highly skewed and most papers are never cited at all. The citation databases are full of inaccuracies, requiring data to be cleaned, which is expensive. Indicators reveal past performance and do not look at present research. Certain fields (such as arts and humanities) and non-English language publications are poorly covered. For these reasons there are initiatives in some countries, for example, Denmark, Australia and by some universities to establish repositories of publications which will be more suitable for performance evaluation. Other indicators used in evaluation include research income from private sources, and numbers of PhDs, spin-off companies and patents, mimicking the indicators constructed for whole country analysis.

One example of a mechanism is the imposition by Ministries and funding agencies of evaluations of all research performers in a certain category: a national research evaluation system. Here, the evaluating organisation imposes a regime upon the research performers who are dependent upon them for funding.

---

UK Research Excellence Framework – a long-established national research organisation evaluation system with funding consequences

The United Kingdom undertakes a national level evaluation of the quality of research in higher education institutes approximately every five years. It is undertaken by the Higher Education Funding Councils (HEFCE) of the UK and employs peer review. HEFCE is an agency (intermediary) which reports to the parent Ministry for Business, Innovation and Skills. The results directly inform the selective distribution of block funding for research to universities in order to concentrate research funding in the best-performing organisations. It also generates benchmarking information for subject areas and of overall institutional performance. It provides accountability to ministers for the public investment in research and allows the funding agency to have considerable steering powers.

An assessment of research quality has been running in the UK since the early 1980s and has taken place six times since 1986, with the most recent evaluation taking place in 2008. The Research Assessment Exercise (RAE), now called the Research Excellence Framework (REF) will be conducted for the first time in 2014. Sub-panels of experts working under the guidance of four main panels will evaluate research under 36 units of assessment (broadly equating to disciplines and subject areas). The quality of selected research outputs, and the vitality of the research environment are judged. The REF will place greater emphasis on the impact of research. This is in response to concerns that the RAE was focused on academic quality at the expense of rewarding user-oriented and more applied research. It will also be informed by indicators, including citation data, for sub-panels which wish to use them, which was not the case previously. A plan to replace peer panels with indicators was abandoned after pilot studies found citation data to be too problematic. HEFCE take charge of specifying all the evaluation procedures for universities and the peer panels to follow, and for ensuring proper execution of the evaluation.

The underlying aim of national research evaluation in the UK has been to divert funding towards the best performing researchers and remove it from the weak research performers. It has done this in the absence of considering national priorities or needs for particular subject areas. For each exercise, the funding formula is not known in advance, but it has become increasingly selective, such that research which is deemed only of national level excellence no longer receives any block grant. Moreover, the definitions of quality have become more and more stringent.

5. The Higher Education Funding Council for England is the largest of the four funding councils (England, Wales, Scotland, Northern Ireland) takes the lead in this evaluation.

Higher Education Institutes and Research Institutes in the Netherlands must show that the research they undertake meets required standards and external accountability obligations. Academic research has been assessed systematically in the Netherlands on a regular basis since the early 1990s. Periodic evaluations usually cover research performed over the previous six years. The Dutch research evaluation system has strong elements of internal organisational learning and only a very weak steering effect as it is not used to re-direct resources.

There are two national protocols employed to assess research. The Standard Evaluation Protocol (2009) is employed by the Association of Universities in the Netherlands (VSNU), together with the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands Organisation for Scientific Research (NWO). The QA System for Research at Universities of Applied Sciences (2009) is used by the HBO-raad. These protocols cover the ex-post evaluation of research performed by a group or institute. In both cases the mechanism is one of self-evaluation and emphasises the learning experience of the assessment through inviting research groups and university management to use it to develop their strategy. The researchers in the group draw up the evaluation themselves and this is assessed by an evaluation committee based on the self-evaluation and interviews with the staff involved.

The objective of this research evaluation system is to encourage organisations to improve their research and management of research through the self-reflection process with a low level of external guidance. The protocol can be applied in a flexible way to match the needs of the research group and its parent institution.


8. The Netherlands Association of Universities of Applied Sciences.
Legal mechanisms for research organisation evaluation – the USA

The United States has the largest national public research system among the OECD countries with $283.8 billion in overall R&D spending in 2003. The federal government provides most of the public support for research. In 1993 the USA introduced results-oriented management for all federal agencies through the passing of the Government Performance and Results Act (GPRA). This was a serious effort to bring in stronger evaluation systems across US federal agencies so that the question “what are we getting for the money we are spending?” could be answered. Research in universities (funded via federal agencies) and government laboratories now had to demonstrate performance and results.

However, in the USA we do not see a national research evaluation system which regularly evaluates ex post performance. The majority of resource allocation in the United States is done through project and individual selection (prospective) rather than retrospective. The focus is on academic quality and, in the case of mission-oriented research, on anticipated benefits and programmes’ fit to mission. The need to meet the requirements of the GPRA has resulted in a greater attention to research strategy in order to demonstrate these aspects.

The Federal law defining the evaluation system at the highest level sits alongside stringent individual-level evaluation. Public research organisations (universities and government laboratories) emphasise individual research performance (grant winning and publication profile). Acquiring tenure or promotion in an American university or government laboratory requires demonstrable performance, as endorsed by peers. Some federal laboratories, including the NIH, use external review for personnel evaluation.

Research in government laboratories is subject to a programme evaluation and review, although most US Federal Reserve funding is distributed on a project basis through competitive merit review. Research evaluation in the mission agencies is closely tied to strategic planning. It uses external advisors and is closely aligned to resource allocation processes. Under GPRA, agencies such as the Environmental Protection Agency and the Agricultural Research Services have adopted strong, ambitious strategic goals and developed extensive new systems for aligning laboratory activities to these goals through an internal project proposal and evaluation process.

---

9. NSF (2005), National Patterns of Research and Development Resources, NSF 05-308, Brandon Shackelford: Arlington, VA.
The NSF and NIH (US basic research agencies) have retained an evaluation system that remains largely unchanged from the pre-GPRA baseline. Both agencies have captured the essence of their missions in strategic plans, systematised the gathering of advances and accomplishments for all their funding activities and the NSF uses external panels to judge agency performance. Formal programme evaluation has not been extended further than pre-GPRA. Evaluation through peer review processes to under-write good performance in the management of research (assuring the selection of the best projects and groups) rather than assessing research results separately is regarded as sufficient. Overall, as in many contexts, we see a variety of mechanisms operating at different levels.

**Negotiating the mechanisms**

Evaluation is a social process with the actors having particular interests in the outcomes. In the Netherlands, evaluations of the Dutch universities have re-shaped the relationships between Government departments and researchers as responsibility for the evaluation has been passed back to university management. In the United Kingdom, the format of the Research Assessment Exercise and its successor, the Research Excellence Framework, has been the subject of open enquiries and intense negotiation between the universities and the funding council. The idea of implementing the GPRA in the USA for research funding agencies brought an outcry, but has settled into the negotiated implementation described above.

**Institutions and ideas**

Evaluation of research organisations is shaped by underpinning ideas about the nature and purpose of research. While we may have left behind the idea of the “Social contract for Science” where scientists are allowed to allocate and manage public funding by themselves in return for robust knowledge, in the USA a great deal of research is regarded as basic knowledge generation and is subject to rigorous peer review and not impact assessment. In other contexts, the idea behind evaluation is that of an investment model, that is, research funding is an investment for future economic growth and social well-being and thus research organisations have to demonstrate economic and social impact. Evaluation of government laboratories with a mission takes ideas from R&D portfolio management in mapping projects onto outcomes and targets.

---

The internationalisation of research has resulted in the idea of international excellence and global positioning of public research organisations. The twin rhetorics of impact and international excellence sit uneasily together since international excellence is assessed by peer review and indicators are prestigious publications, while socio-economic impact requires the application of methodologies from economics.\textsuperscript{12}

Some observers assert that evaluation of public research is part of the bundle of ideologies and practices loosely termed “new public management”\textsuperscript{13}. These include notions of accountability, performance measurement, cutting down on the core functions of the state and devolving management to agencies. While evaluation of public research organisations does sit alongside audit, performance indicators and targets in some contexts, many of its features have been adopted across nations without strong NPM reforms as part of the changes in research and the role of research organisations.

Despite periodic calls for it, public debate and shaping of research seems to have little to do with research organisation evaluation, with the possible exception of the Netherlands. An interesting development is the addition of societal relevance of public research to the protocols, which already covered quality, productivity and vitality and feasibility. The self-evaluation protocol assumes that societal relevance is evaluated in the context of the research group’s mission, interaction with stakeholders and results that might be of importance to society. The evaluation committee can meet the stakeholders or can include them in its membership. The self-evaluation requires groups to reflect on which of their research results might be important to society, how they have disseminated the knowledge to societal stakeholders, what interest societal stakeholders have in their work and what impact their research has had.

**Incentives**

The incentives for evaluation naturally vary according to the actor and their position with respect to evaluation. Funding agencies and Ministries need to demonstrate accountability and value for money and to be able to compete for public funding with other areas which have more political relevance, such as health and social care. Incentives for research performing organisations include simple compliance: if they do not co-operate they will not receive their funding. However, beyond this the performers also need to justify their activities and demonstrate their quality and relevance. The reputational effects of evaluation systems are probably just as important as the financial implications.

\textsuperscript{12} e footnote 1.

\textsuperscript{13} Paredeise, C., Reale, E., Bleiklie, I., Ferlie, E. (Eds) (2009), University Governance: Western European Comparative Perspectives Springer.
Evaluation of research organisations in itself produces some perverse incentives as researchers and organisations will always seek to achieve the highest rating possible. The increasing popularity of league tables influences and shapes the behaviour of some types of public sector research organisations, in particular universities. The best known international ranking of universities is the Academic Ranking of World Universities (ARWU) compiled by the Shanghai Jiaotong University. Each ranking system uses different indicators and these are mostly related to research, since it is so difficult to benchmark teaching quality. University league tables are powerful signals for prospective students and funders, national and international partnerships and donors. They distinguish elite from mass institutions. However, few regions can afford the investment required to run a “world class” university.

**Perverse incentives in Research Organisation Evaluation**

Evaluation mechanisms can have unintended effects on academic researchers and the type of work they undertake. These include affecting the choice of publication practices to suit what is considered to be the best strategy for a good outcome. In evaluation systems which count publications there is evidence of spreading material across several papers to achieve an increased publication rate. A commonly raised concern is a move to journal publications in disciplines where other outlets (such as monographs) have traditionally been more important. There is an unintended impact on staff morale in systems where staff can be left out of assessment or labelled as “non-research active”. The status of teaching within the university sector suffers if research is explicitly or implicitly a more valued activity, so that excellent teaching is not rewarded and students feel neglected. Certain types of research are clearly disadvantaged in peer review referring to international excellence, including practice-related and applied research, and research not of international standing but of local or regional significance. There is a possible impact on collaboration, with some assessment systems discouraging joint publication within departments where researchers of multi-authored outputs must decide who takes the credit. There are also impacts on university behaviours as they try to maximise their success by recruiting ‘star’ appointments just prior to assessment date and within departments to form ‘citation clubs’ of colleagues.

---

Scope for policy intervention

We can consider evaluation to be one of the tools available to the policy maker along with foresight and strategic planning. As such, it is a mode of policy intervention and the evidence suggests that public research organisations can be highly sensitive to its dynamics and influence. Both individual researchers and public research organisations may object to the introduction of evaluation, on the basis that it violates academic freedom and wastes time and resources in assembling evaluation information and taking part in reviews. The ability of governments to implement evaluation will depend upon the strength of the incentives, the strength of research élites and the independence of the research organisations. Even where governments have considerable power (in budgetary terms) they still need to negotiate the terms of evaluation with research organisations, and they in turn need to communicate with individuals and groups to induce cooperation. Outside parties, such as industry and other research users (for example, patient groups, environmental groups) may also have influence over whether and how public research organisations are evaluated.

Assessment

A substantial investment is made in evaluation, particularly when the indirect costs of the time devoted by management and peers in delivering evaluation is taken into account. It is reasonable to ask what are the outcomes and uptake of evaluation and whether it makes a difference to policy. The impact of evaluation is often indirect, for example, leading to behavioural changes within research organisations such as strategic planning. The impact of evaluation on research funders is also often indirect because it is not politically or administratively simple to shut down public institutions as a result of a poor evaluation. Instead there might be a reorganisation or a change of mission.

Overall, research organisation evaluation seems to be increasing complex, as we can see with the evolutions of the national research evaluation systems requiring elaborate peer review processes, which are notoriously difficult to implement in a consistent way. There may be diminishing returns in terms of improved performance or re-distribution of resources. Researchers and their managers are adept at modifying their behaviour to get the best possible evaluation result rather than the best quality research. The costs of evaluation are contested, but seem to be increasing, and so it is likely that the cost-effectiveness of evaluation is in doubt. The opportunities which evaluation brings for learning, re-positioning and reputation-enhancing should be emphasised over simple rankings and funding allocation.
Further resources

Academic Ranking of World Universities, http://www.arwu.org/index.jsp


Butler, L. (2009), ERA (Excellence in Research for Australia): Update, http://www.kcl.ac.uk/content/1/c6/07/31/59/LindaButler.ppt#258,22,Lessons for REF?


OECD-Norway Workshop on Performance-based Funding for Public Research in Tertiary Education Institutions, presentations available at http://www.oecd.org/document/55/0,3343,en_2649_34293_45711095_1,1_1,1,00.html

