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

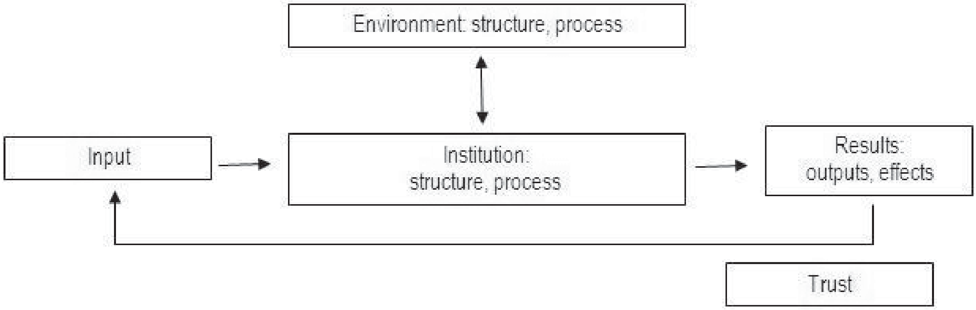
Performance-based management has become widespread in the public and not-for-profit sectors. A variety of terms are used in the literature on this topic. Besides performance-based management there is results management, result-based management, managing for results, managing for outcomes, outcome-focused 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.

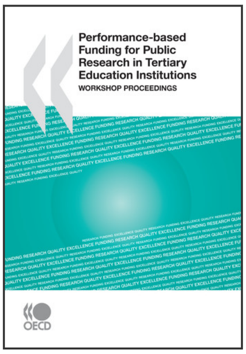
Thiel, S. van and F.L. Leeuw (2002), "The Performance Paradox in the Public Sector", *Public Performance & Management Review*, Vol. 25, No. 3, pp. 267-281.

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.



From:

Performance-based Funding for Public Research in Tertiary Education Institutions

Workshop Proceedings

Access the complete publication at:

https://doi.org/10.1787/9789264094611-en

Please cite this chapter as:

Hansen, Hanne Foss (2010), "Performance indicators used in performance-based research funding systems", in OECD, *Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings*, OECD Publishing, Paris.

DOI: https://doi.org/10.1787/9789264094611-5-en