Department of Finance
Ministère des Finances

Working Paper
Document de travail

Methods of Evaluating University Research Around the World

by

Aled ab Iorwerth

Abiorwerth.aled@fin.gc.ca

Working Paper 2005-04

March 2005

.

Abstract

This paper looks at the increasing trend around the world to evaluate the quantity and quality of universities' research output. Evaluations can provide a useful role in clarifying the performance of university research, and give incentives to increase research output and quality because of increased competition between institutions. Methods of evaluating university research output deployed across countries are discussed.

Résumé

Il y a une tendance croissante dans le monde à évaluer la quantité et la qualité de la recherche universitaire. Les évaluations clarifient la performance des universités, et peuvent fournir des incitations à augmenter la qualité et la quantité de la recherche en augmentant la concurrence entre les institutions. L'auteur examine les différentes approches couramment utilisées dans le monde pour évaluer la recherche universitaire.

Table of Contents

List of Figures

List of Tables

1. Introduction

It is widely believed that there are significant social benefits from increasing scientific research output. Because of the potential to increase economic growth, many governments around the world have increased resources devoted to university research. But the ability of increased resources to increase research output depends on the productivity of the research system. A key step towards increasing research productivity is to measure both the quality and the quantity of universities' research output. The importance of raising research productivity may explain why evaluation of university research outputs has been an increasing trend around the world over the last twenty years.

The objective of this note is to provide some of the underlying principles employed in evaluation and to highlight, in a stylized way, how evaluation is undertaken. Although the principal goal is to examine evaluation systems, this discussion takes place within a wider context of increasing interaction between scientists and society, a context that will be discussed briefly in this introduction.

The traditional view of 'science' was an activity 'based upon divine cognition done by individual researchers.'[1] The physicist John Ziman has called this approach 'the Legend': the 'mythological view of science as the product of the entirely rational and selfless efforts of dispassionate humans'.[2] A further tradition was that scientists themselves set the direction of science with the goal of expanding the research frontier. There is of course open discourse within science but science itself was essentially a closed system. This 'separateness' of science is changing.

Science takes place within a social context.[3] Universities, government, industry or society at large define the social context, and changes in these institutions can drive changes in the organization of science. This note does not examine the behaviour of individual scientists but instead concentrates on the institutional structure in which they operate.

The interaction between society and science works both ways. Not only do institutions' decisions impact on science but science itself impacts more heavily on society and institutions, and not only in terms of resources. Unquestioned trust in scientific knowledge is eroding. In his BBC Reith Lectures, Anthony Giddens remarked

> Our relationship to science and technology today is different from that characteristic of earlier times. … Most of us—including government authorities and politicians—have, and have to have, a much more active or engaged relationship with science and technology than used to be the case. We cannot simply 'accept' the findings which scientists produce, if only because scientists so frequently disagree with one another. …We do not currently possess institutions which allow us to monitor technological change, nationally or globally. The BSE debacle in Britain and elsewhere might have been avoided if a public dialogue had been established about technological change and its problematic consequences. More public means of engaging with science and

[1] Hemlin and Wenneberg (2002).
[2] From McConnell and March (2001).
[3] The intention here is not to suggest that there is no such thing as 'pure rational' science but only to note that scientists do not operate in a world apart. For an examination of the fight between scientists and sociologists, see Sokal and Bricmont (1998).

technology wouldn't do away with the quandary of scaremongering versus cover-ups, but might allow us to reduce some of its more damaging consequences.[4]

With the impact of science on society the closure of the scientific system is increasingly untenable.

Increasing interaction with society has led academics that study science to suggest that there are changes in both science and the production of knowledge.[5] In Figure 1 this change is depicted as a movement from mode 1 to mode 2.[6] This change causes new and fundamental challenges of defining what is new scientific knowledge and of establishing its trustworthiness. These new questions challenge the traditional methods of science because, within the closed world of science, traditional methods only satisfy scientists. Scientists now need to persuade society that they produce social robust knowledge.

Figure 1: Knowledge Production

Mode 1	Mode 2
Problems set and solved in the context of the (academic) concerns of the research community	Problems set and solved in the context of application
Disciplinary	Transdisciplinary
Homogeneous	Heterogeneous
Hierarchical, tending to preserve existing forms of organization	Heterarchical, involving more transient forms of organization
Internal quality control	Quality control is more socially accountable

Source: pp. 12, Arnold et al. (2001)

The increased challenge for science also leads to challenges for institutions that foster science:

Efficient Allocation of Scarce Resources

- To develop mechanisms so that support for research can be prioritized within a discipline, within a university and within or by governments. This goal is key for policy makers responsible for allocating resources;

Reaching out to Wider Society

- To develop evidence to the wider public of the social and economic benefits of supporting research.

To a certain extent, the introduction of evaluation can be a first step to meeting these challenges. The information provided by the evaluations themselves can help. But, if evaluation processes are established with the cooperation of academics, hence increasing their legitimacy, then these processes can begin to establish more interaction between academics and society.

Evaluations can achieve the first goal in two ways: 1) by providing incentives for research institutions to emphasize their mission of increasing research, and 2) by providing at least some information on practices that could lead to higher research outputs. Evaluations give the

[4] pp. 31-35, Giddens (2003).
[5] It can be argued how true this change has been in all of history. But this change seems to be an adequate representation for a change from 'Cold-War' science to the structure of science today.
[6] For a wider discussion, see Gibbons et al. (1994) and Nowotny et al. (2001).

opportunity for institutions to learn and also give information on how successful is the research system at producing outputs. Rather than being aimed at the academics themselves, these evaluations should therefore be directed more at those responsible for program implementation (so that they can reallocate resources, reassign responsibilities, modify objectives).[7]

For policy makers, the challenge, put somewhat starkly, is the following:

> The federal government needs to measure whether its R&D investments are effective. We can rarely show what our R&D investments have produced, and we do not link information about performance to our decisions about funding. Without this information, decisions about programs tend to be made on the basis of anecdotes, last year's funding level, and the political clout of local interest groups.[8]

On the second goal above, evaluation can influence the responses of outsiders.[9] With increasing demands on the public purse, it may become increasingly necessary for universities to legitimize expenditure on research. Through accountability universities can build social capital. As the Association of Universities and Colleges of Canada (AUCC (1995)) noted as far back as 1995,

> In this period of fiscal restraint, the emphasis is on 'value for money', or ensuring that funds invested in public education are well spent. Increased demands for accountability throughout all sectors of society have led to an interest in performance and outcome measures … Universities themselves recognize the need to adapt to the fiscal realities and seek new information on which to base decisions concerning efficiency, effectiveness and productivity.

It would be in universities self-interest to demonstrate their efficiency and effectiveness by publishing output measures that are more comprehensible and meaningful to the public.[10] In the absence of such reassurance, the OECD notes how claims that basic research leads to appreciable long-term benefits may lead to disillusionment by the public when visible benefits are not forthcoming.[11] For example, the Australian Academy of Science (AAS (2003)) notes how, "[i]n the UK, the RAE has not only improved the quality of research outcomes, but has contributed significantly to public and government awareness of the excellence of its scientific enterprise."

The Comité National d'Évaluation de la Recherche (CNER(2003))) in a report to the French President reports,

> Au fur et à mesure que les méthodes d'évaluation sont devenues systématiques et professionnelles, de même de la gestion de la recherche, dans une grande mesure pour répondre aux besoins de l'évaluation, il est devenu nécessaire de préparer des plans d'action stratégiques pour améliorer les performances. C'est le cas de la plupart des établissements aux Pays-bas. En Finlande, la concentration de l'évaluation et des ressources financières depuis les années 70 sur des centres d'excellence a permis le développement de la recherche dans de domaines définis comme stratégiques. Par la même occasion, dans ces pays et ailleurs, l'évaluation semble avoir renforcé la

---

[7] This call is essentially for improved management, http://www.remap.dk/departments/mpp/remap/publications/science.pdf
[8] pp. 43 OMB (2002).
[9] Feller (2000)
[10] pp. 80 in OECD (1999).
[11] pp. 81, OECD (1999).

> légitimité de l‘investissement dans la recherche et la confiance du public dans la science et la technologie.[12]

The logic of the paper proceeds as follows. Assuming that the government is interested in increasing research outcomes—quantity and quality—the paper first explores whether there are different outcomes across countries, which is indeed found. The factors that account for these international differences are numerous however, expenditures only being the most obvious. This paper concentrates on how evaluation systems differ around the world. Many countries with high research performance either have or are moving towards evaluation systems that assess research output.

The paper then discusses the methods of research assessment employed and their various pros and cons. The experience of the U.K. is then discussed, as there is some evidence that the U.K. research system is very productive. Evaluation methods have been employed around the world and the evolution of implementing these evaluations is discussed, concentrating on the Netherlands, Denmark and Australia. Introducing such systems is a learning process and in all countries there has been a shift from the system as it was originally conceived. Many other countries are moving towards or evaluating research evaluation systems.

The conclusions of this paper closely mirror two recommendations made in a report by the Standing Committee on Industry, Science and Technology in the Parliament of Canada:

- That the Government of Canada require the granting agencies to engage in more regular internal review of their own programs and practices (including peer review), and to periodically examine decision-making processes at other Canadian and foreign agencies to ensure that best practices for the allocation of research funds are in place. The results of these internal evaluations should be easily accessible to the research community and general publics.

- That the Government of Canada ensure that the federal granting agencies take steps to better measure and report on the outcomes, and where possible, impacts of their research programs for the benefit of the general public.[13]

[12] pp. 29 CNER (2003).
[13] Recommendations 10 and 11 of Parliament of Canada (2002). For the discussion in the report of the difficulties in the peer review system and the basis for the recommendations, see their Chapter 5.

2. How does research output vary across the world?

This section gives some background on research outputs around the world. [14] There is no unique way of comparing research output so this section will look at a range of indicators that may vary across different sources. Because total research output may be strongly influenced by the size of a country, this section will concentrate on measures of 'productivity': research output per unit of resource. Figure 2 captures, at a cursory level, some stylized facts of the effectiveness of research institutions across OECD countries. Small central and northern European countries are research powerhouses in terms of publication output per population. Also entering this group are Australia and the U.K., which is noticeable in being a relatively large country but has maintained its prolific output. Canada does relatively well, being eighth in the OECD group.

A problem with using total scientific papers as an output measure is that it does not control for quality differences. With a belief that more highly cited papers are of better quality, citation rates are usually used to control for quality. The proportion of highly cited papers is shown in the upper quadrant in Figure 4, which is taken from a European Commission publication (with no Canadian data). These same countries, with the addition of the U.S., produce high quality papers. Large European countries, other than the U.K., do poorly. A constant feature of the data is that, as far as publication output goes, France and Germany do not do well. There are many possible explanations for this result: low propensity to publish in English[15]; universities perform a relatively small share of research; or academic researchers in France and Germany may interact more with industry.

Another possible measure of research output is shown in Figure 3, which shows the percentage of researchers in the workplace. Researchers are an input into the research process, but they are also an 'output' of research universities. Northern European countries again predominate.

Summarizing some of the results above, Table 1 shows the top ten countries by different rankings of research output. The first column shows that larger countries obviously produce more research output. The ranking generally matches the ranking of GDP. The exceptions are the U.K., which was fourth in terms of GDP but third in terms of total research output, and Australia and the Netherlands who ranked eleventh and twelfth for GDP in 1999.[16] In terms of intensity, productivity and quality—the next columns in Table 1—it is also clear that many small open economies are much more efficient at producing high-quality research than some large countries.

Figure 5 looks at the data on scientific publications over time. A pattern that emerges is the growing importance of some parts of Europe and Asian countries as sources of scientific publications. The number of publications in Taiwan went up over six-fold and in South Korea by a factor of ten (both not shown) over 1986-2000. The U.S. shows a relative decline but its base is of course much larger than other countries. Canada also shows a marked decline whereas the U.K. has been roughly constant since about 1989.

---

[14] It should also be borne in mind that publishing research papers should not necessarily be the sole objective of government policy (increasing the supply of skilled workers is also important). Consequently concentrating on learning lessons only from Sweden's proclivity to publish may lead to the wrong lessons being learned since Sweden apparently has a low rate of commercializing its research output. See Henrekson and Rosenberg (1998).

[15] Sweden, Switzerland and the Netherlands, being smaller countries, moved to publishing in English much earlier.

[16] These rankings use GDP in purchasing power parity currencies for 1999 for OECD countries. Mexico was seventh in 1999 but had fallen to eighth in 2002.

Figure 6 shows whether countries are ranked in the top three countries over 1993-2002 for citations in any of nine research fields.[17] The U.S. is always in the top three in all nine disciplines and the U.K. is in the top three in seven disciplines for all years but one. For reasons that are unclear, Canada has dropped from having high quality research in four fields to being in the top three in only one field in 2002.[18]

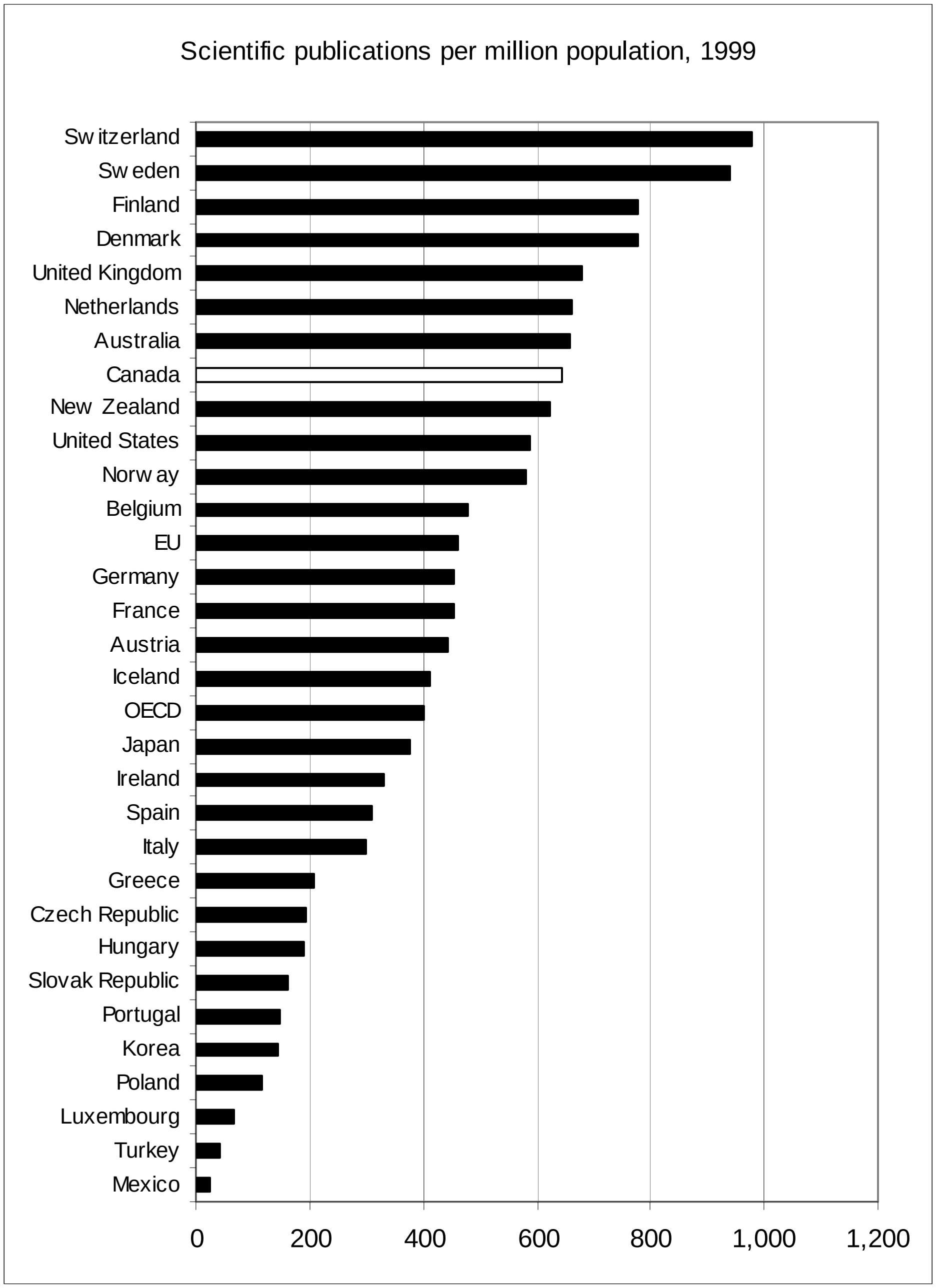
Figure 7 shows a 'top-5' ranking of countries (from a different source) in each of seven fields and the position of Canada. Canada is only in the top five in clinical medicine and social sciences. It is noticeable that the U.S. and the U.K. are the only large countries regularly in the top five. Many small countries figure prominently and consistently.

The general impression from the previous discussion is that Canada's research performance is fairly good on average. However, the data on citations suggest that Canada is not producing as much 'world-class' research as it could. The performance of countries like Finland, Sweden and Switzerland and the U.K. suggest that more research output could be obtained.

---

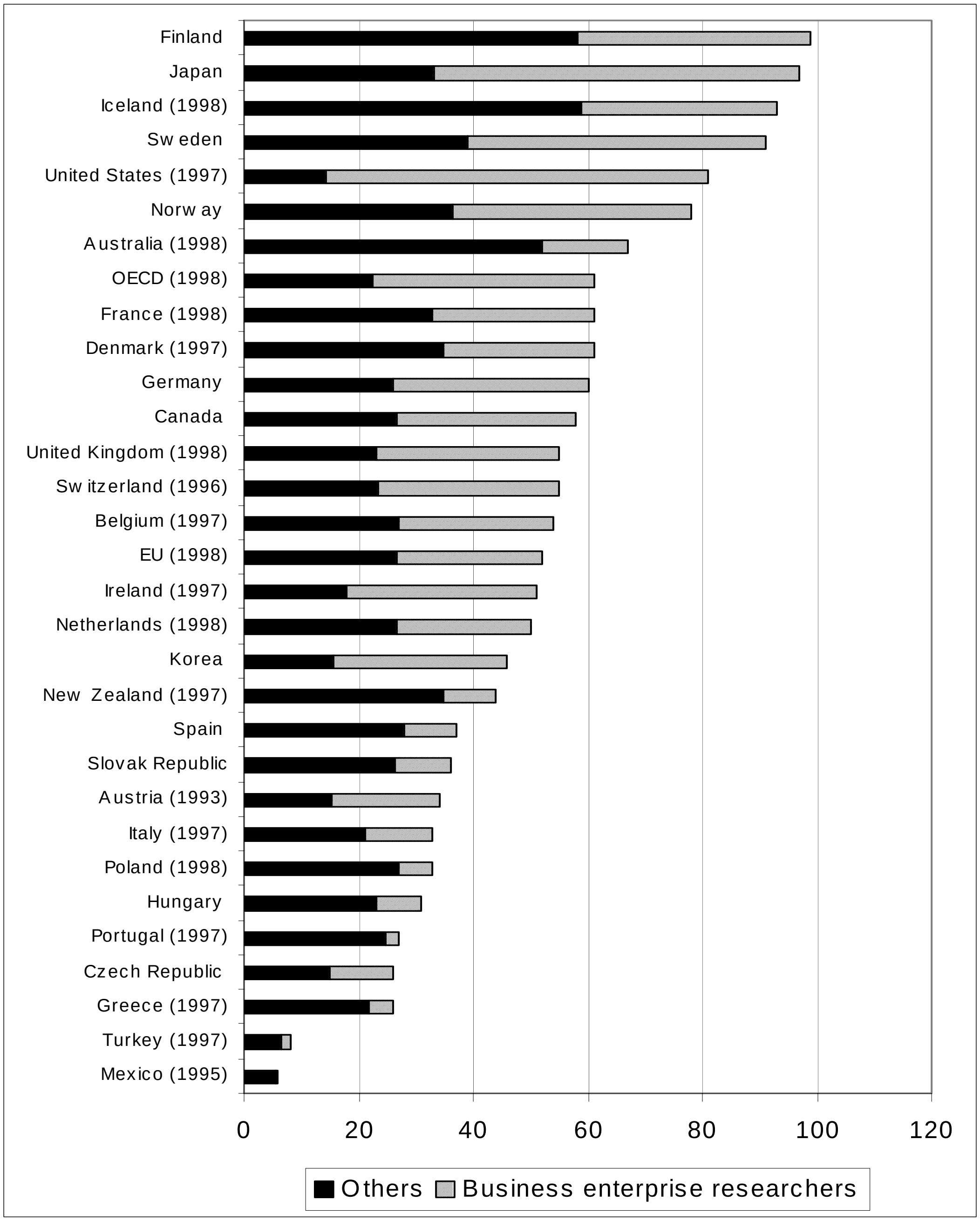
[17] The fields are: clinical sciences, pre-clinical and health-related sciences, biological sciences, environmental sciences, mathematics, physical sciences, engineering, social sciences and business.
[18] Canada is third in citations in the social sciences, having dropped to fourth in business in 2002.

Figure 2: Scientific publications per million population, 1999



Source: OECD (2001a).

Figure 3: Researchers per 10,000 labour force



Source: OECD (2001b)

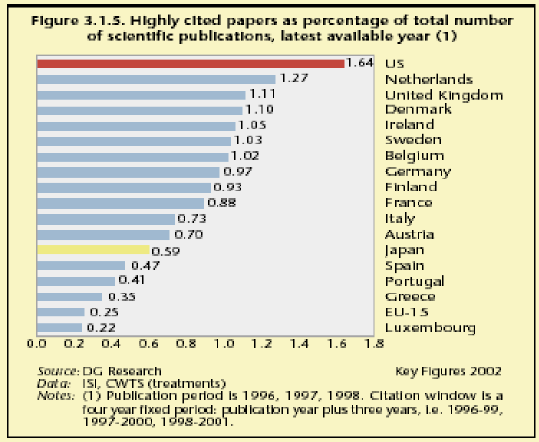
Figure 4: Citation Data

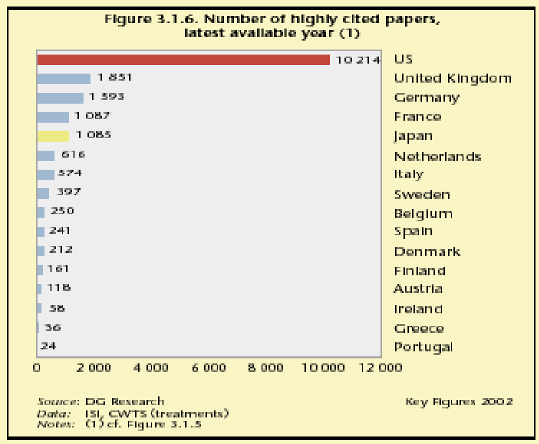
Highly cited papers

High citation rates are used as a proxy of the importance of research. Only a small fraction of the overall scientific output is highly cited. However, the number of papers and the analysis of the particular fields in which these papers occur, give some important information to science policy-makers, and signal important research to scientific communities worldwide.

For Key Figures 2002 the length of the citation window and the retrieving method have been amended (cf. Annex II).

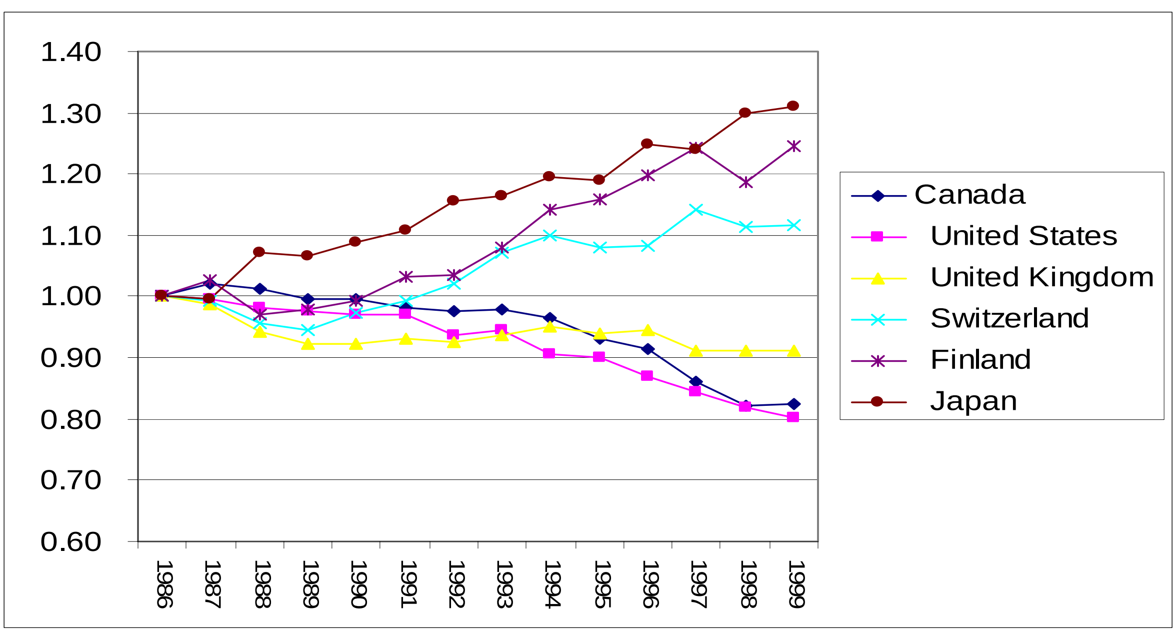
When it comes to highly cited papers, all EU countries lag behind the US. However, there are several countries with shares above the world average. The Netherlands, UK, Denmark, Ireland, Sweden and Belgium do far better than the world average (1.0), while Italy, Austria, Spain, Portugal and Greece achieve only below average shares. The absolute numbers reflect the size of the publishing country. It should be noted that the calculations are based on a full counting method. The number of highly cited papers is not additive but international co-publications are counted fully for two or more countries. It would be interesting to estimate the percentage of European involvement in highly cited US publications and vice versa.





Source: Taken from pp. 48, EC(2002).

Figure 5: Trends in the share of scientific publication: 1986-1999



Source: NSF(2002).

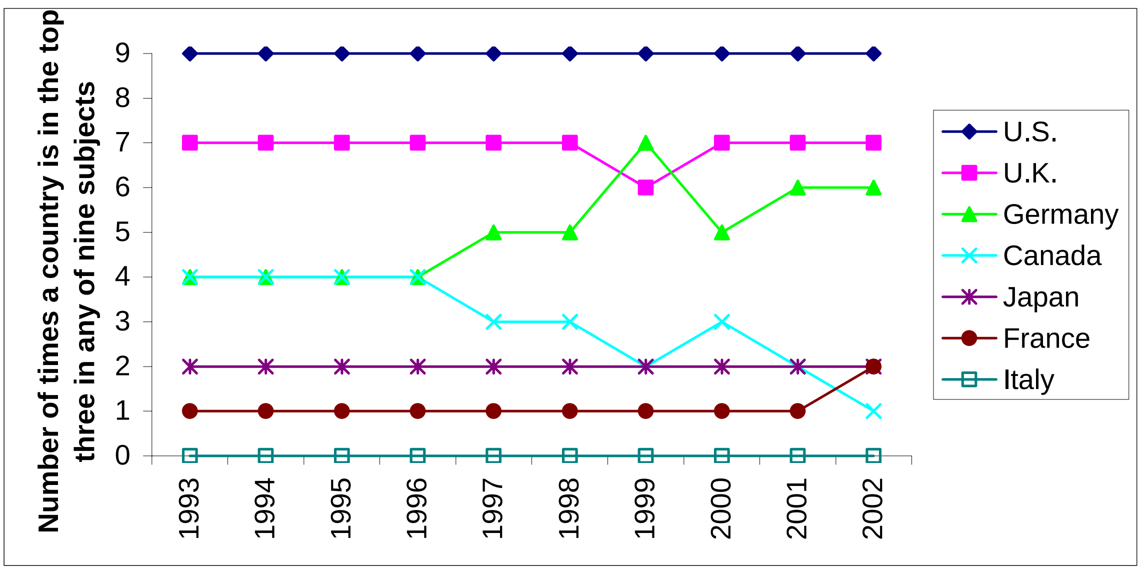
Table 1: Ranking by measures of research output

	Country share of scientific publications in the OECD total, 1999	Scientific publications per population, 1999	Scientific publications per higher education researcher, 1999†	Scientific publications per total researcher, 1999†	Rank by citations per paper for the top 1% of papers (to 2000)
1	U.S.	Switzerland	U.S.	Switzerland°	Switzerland
2	Japan	Sweden	Netherlands	New Zealand	U.S.
3	U.K.	Finland	U.K.*	Italy	Netherlands
4	Germany	Denmark	Switzerland*	Netherlands	Denmark
5	France	U.K.	Denmark	U.K.*	Sweden
6	**Canada**	Netherlands	Italy	Denmark	U.K.
7	Italy	Australia	Austria*	**Canada**	**Canada**
8	Australia	**Canada**	Czech Rep.	Sweden	Finland
9	Spain	New Zealand	**Canada**	Spain	Belgium
10	Netherlands	United States	Sweden	Australia*	Germany

Sources: OECD(2001), Evidence (2003). * Data for 1998 ° Data for 2000

† Scientific publications are more likely to be produced by higher education researchers but some government and large firms re undertake scientific research. Because of the U.S.'s large number of private-sector researchers, its ranking in scientific publications per total researcher is lowered to 21st.

Figure 6: Leading nations for ranked citation volume in nine research fields



Source: Thomoson ISI, reported in Evidence(2003)

Figure 7: Ranking of country according to impact, 1998-2002

	Engineering, computing and technology	Physical, Chemical and Earth Science	Agriculture, Biology and Environmental Sciences	Life Sciences	Clinical Medicine	Social and Behavioural Sciences	Arts and Humanities
1	Switzerland	U.S.	Netherlands	U.S.	U.S.	US.	Netherlands
2	U.S.	Switzerland	Denmark	Switzerland	Netherlands	Netherlands	U.S.
3	Israel	Denmark	U.K.	Netherlands	Finland	U.K.	U.K.
4	Denmark	Netherlands	U.S.	U.K.	Sweden	**Canada**	Finland
5	Netherlands	Germany	Sweden	Finland	**Canada**	Finland	Israel
Canada's Position	11	8	9	6	5	4	6

Source: Scientometrics Scoreboard, CEST, Switzerland

3. Is University research important?

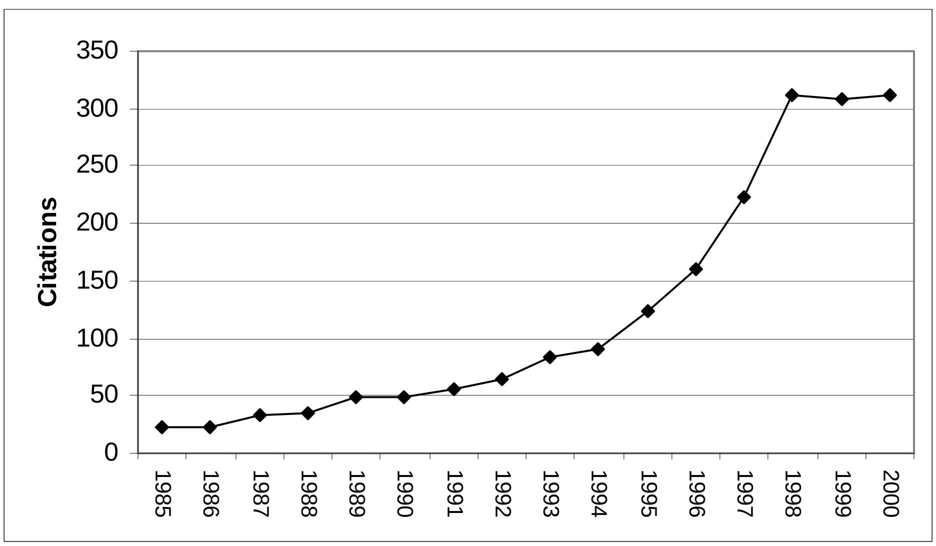
A valid question to ask at this point is whether society should care about either the quality or quantity of academic publications, or even science writ large?

The potential importance of developments in the pattern of publication is shown in Figure 8, which shows that 'science' is becoming more important for developing new commercial products. Figure 8 shows that references to scientific articles, not patents, increased dramatically over the 90s.

The left panel of Figure 9 shows that universities are becoming a more important source of patents. However the importance or significance of universities' patents is declining on average suggesting that there are decreasing returns to patenting.[19] This admittedly cursory analysis suggests that policy should certainly preserve and reinforce the core mandate of universities to produce high quality basic research.

One should bear in mind that 2 per cent of the world's research institutions (broadly defined) account for 69 per cent of publications and 79 per cent of world citations.

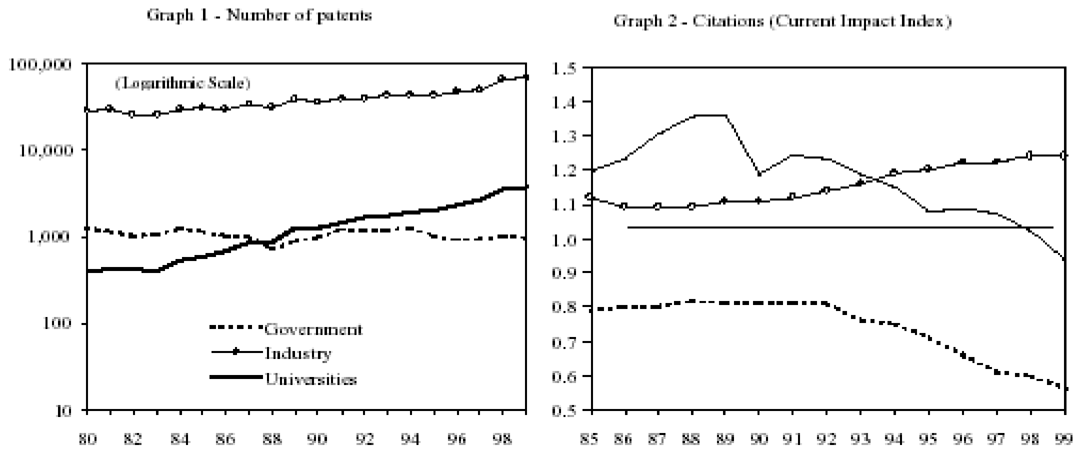
Figure 8: Citations in U.S. patents to scientific and technical articles



Source: NSF(2002). Citations include all references to scientific articles. Citation counts are on the basis of a twelve-year period with a three-year lag; for example 2000 citations are references of U.S. patents issued in 2000 to articles that were published 1986–97. Changed U.S. Patent & Trademark Office procedures, greater ease of locating scientific articles, and greater incentive to cite them may have contributed to some of these increases.

[19] See Henderson, Jaffe and Trajtenberg (1998) and results in Hicks, Breitzman, Olivastro and Hamilton (2001).

Figure 9: The rise in university patenting



Source: Hicks et al. (2001).

.

4. Policy questions. Can we learn from other countries?

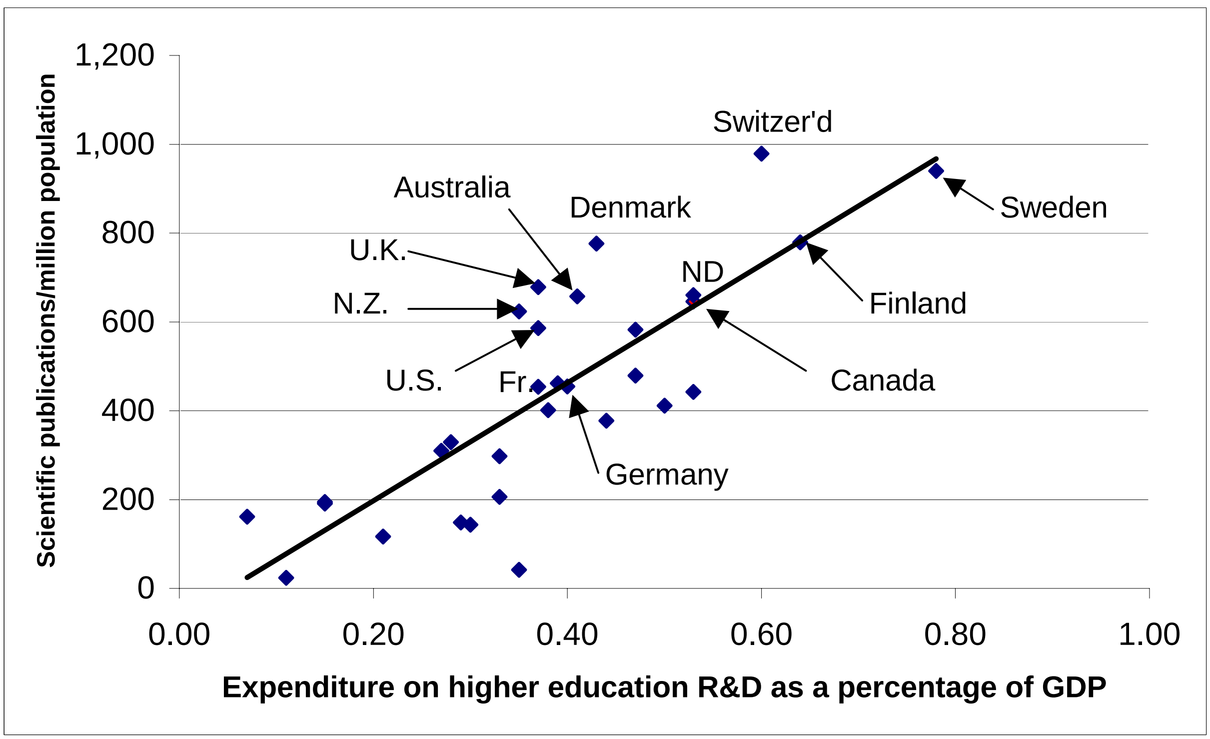
Even if one accepts that 'science' is important for future growth, the following questions need to be asked:[20]

1) How much money should the state invest in R&D?
2) What is the best place to invest the money along the spectrum from basic research to innovation and diffusion?
3) Is a better return available from investing in some sectors, branches or clusters than in others?

A prerequisite for answering these types of questions is that information is available on the impact of research initiatives. At the moment, these questions cannot be answered satisfactorily.

Figure 12 displays a measure of research output—scientific publications per million population—against a measure of research input—money expended as a proportion of GDP. One could, crudely, expect the publication rate to increase, in some fashion, as resources spent increased.[21] In Figure 10 the linear trend line shows how output could increase with more resources.[22]

Figure 10: A measure of research output against research input, OECD countries 1999



Source: OECD (2001a, 2003).

[20] Erik Arnold and Ken Guy, 1997, "Technology Diffusion Programmes and the Challenge for Evaluation", pp. 65-87 in OECD(1997).

[21] Perhaps a more realistic relation is one of diminishing returns: as more resources are devoted within a country, research output does not increase as rapidly.

[22] A simple univariate regression of publications on expenditure yields an equation of *PUB=-68.1+1327.79EXP* with an R^2 of 0.65.

With a linear trend some countries get the average return (according to this measure) from their inputs: France, Germany, Canada, Sweden and Finland. However, some countries seem to get more publications per unit resource than others: Switzerland, Denmark, Australia, New Zealand, the U.K. and the U.S.[23]

It should also be borne in mind that publication is only one possible output indicators. There is some evidence that increased resources will increase the number of publications but quality may suffer as resources increase. Using data on U.S. universities, Abigail Payne and Aloysius Siow (2003) find that an increase of $1 million would result in 10 more articles and 0.2 more patents, but has no effect on citations per article. These results could suggest that increased resources produces more output but not necessarily excellence. The top panel of Figure 4 gives a clearer idea of excellence.

Looking at some of the previous figures, it is clear that resources are important in determining research output. But it is also clear that other factors play a role as well.

How do countries differ?

Research output is a function of the resources expended and the institutional structure (or the microeconomics) in which research is undertaken. Countries differ in the quantity and quality of output that they produce. The obvious next question is are there institutional features that explain these differences in outcomes, independent of how much money is spent?

The first part of this section takes a macro look at how research is funded around the world. In particular, it draws a distinction between systems that use some indicators of performance and those that do not. No attention is paid at the moment to the details of these performance systems. The next part of this section looks at other possible explanatory factors.

a) Organization and Research Performance

A central theme of this paper is the methods by which universities are funded and the (possibly separate) existence of evaluation systems impact on incentives to undertake research. These incentive changes can lead to improved efficiency of research. Consequently, instituting funding and/or evaluation changes are the principal ways that will be examined here as means of altering research output.

To a certain extent all funding is based on some evaluation since all funding for universities are linked to some yardstick. However, such yardsticks may not include research output. The extent to which research is evaluated therefore varies across countries depending on how universities are funded.

The first column of Table 2 shows a full range of how universities receive money under the various systems in use around the world. Each type of funding system has an associated evaluation system. These systems are of course not mutually exclusive. Traditionally, universities received a block grant and research funding was a residual of this block grant. Other

[23] If there are diminishing returns to the absolute dollars spent on research—a smaller increase in publication for each additional dollar—then the performance of some of the larger countries is even more creditable. If there are diminishing returns to scale to resources spent within a country then the performance of Finland and Sweden becomes more creditable.

traditional methods are given in the next two rows. University funds in the Netherlands are based on what a university has received in the past, and universities in Austria and France obtain funds based on negotiations with central agencies.

With the importance of universities as sources of science, other types of funding have also become common over the last decades. Direct funding of research by granting councils is usually based on a competitive process where peers evaluate the research proposals. Government can also earmark research funding. The process in this case may also be competitive although the competition may be played out in the political field.

As well as block funding (the first type of funding), Canada currently uses, more or less, these last two methods. Provinces provide direct funding, which is largely conditioned on measures of university size.[24]

However, an increasingly important source of funding around the world is funding that is in some way based on evaluation of research outputs. Many northern European countries have moved to systems that base funding implicitly or explicitly on research performance.

The proportion of research that is evaluated and the manner in which it is evaluated is therefore a function of how universities receive their resources. For example, the U.K. uses the last three methods in Table 2 and will therefore likely have, *ex ante,* a higher proportion of research evaluated for quality than Germany, which largely uses block grants. Increasing one method of funding will therefore increase the associated evaluation based on that type of funding.

An important example of the impact of the institutional structure is the evolution of research funding in Finland.[25] Early in the 1990s, Finland made science and technology policy a strategic core in the formulation of policy. For example, budget cuts forced upon Finland as a result of a general economic crisis were not applied to research. The result of Finnish policy is that Finland is seen as a success story of strategic science and technology policy.

This science and technology policy was also applied to universities. Prior to reforms, Finnish universities had existed in a world of their own. They were very skeptical about increased interaction between universities and business, for example. But during the 1990s, the government changed how universities were funded through increasing the proportion of funds that they allocated by a competitive process. Between 1990 and 1996, competitive funds' share of the government transfer increased from 42 per cent to 52 per cent.[26] The proportion of universities' budget allocated by competitive and other external funding increased from 33 per cent to 47 per cent. Clearly as the Finnish system moved towards a competitive process as the proportion of research that was evaluated increased. As discussed in the previous section, this change does not seem to have damaged Finland's research performance. Is should also be noted that, of course, the Finish granting councils are themselves evaluated (see below and Appendix 2).

---

[24] Alberta has introduced performance indicators to determine the size of the block grants. Some indicators refer to success in obtaining research council grants.

[25] The following two paragraphs draw pm Chapter 2 of Government of Denmark (2003).

[26] In comparison, total federal funds in Canada constituted 13.3% of total government funds, http://www.statcan.ca/english/Pgdb/educ47c.htm, which is roughly comparable to other countries. In the U.K. research council funding was 9 %, in Denmark 20 % and in the U.S. about 30 % in 1999. Jongbloed and Vossensteynm (2001).

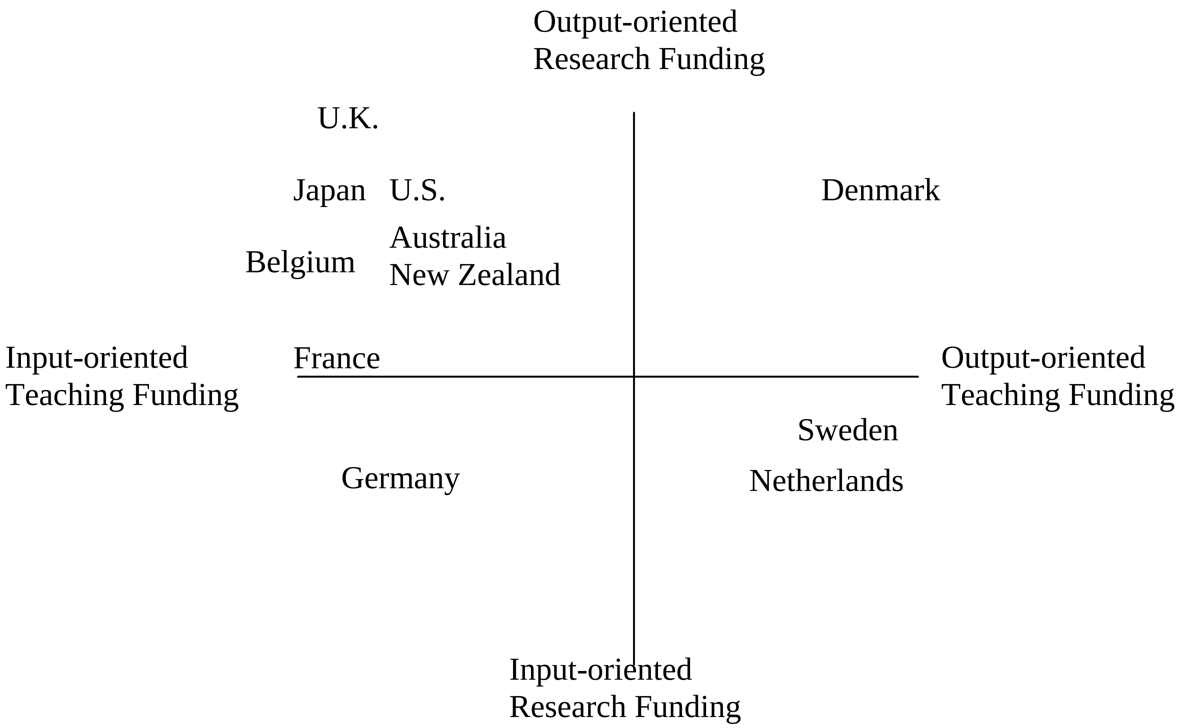
Table 2: Funding and evaluation in sample countries

Method of transferring resources	Evaluation undertaken	Example countries
Core funding or block grant funding	No evaluation of research.	a) Block grant wholly based on number of students: Germany, Italy b) Research-field evaluation Norway, Sweden c) Some allocation of funds according to a performance measure: Finland, Denmark
Historical	No explicit evaluation.	Netherlands (Evaluation takes place but is not linked to funding)
Negotiating	May involve quality assessment	Involving evaluation of research: France No evaluation: Austria
Direct funding of research projects by granting councils	Peer review	Most countries
Contracts	Output	Most countries
Research Performance Indicators	Based on evaluation of research quality	Numerical indicators: Australia, Poland Informed peer review: U.K., Hong Kong

Output-oriented funding can also be introduced for teaching. In Figure 11, one can see the variety of systems that have been put in place around the world. Generally however, there is an increasing concentration of countries that are in the northwest quadrant: there is increasing amount of resources that are allocated on the research output of universities. But only a few countries have moved to have an output-based system for the teaching of students.

Table 3 shows some of the advantages and disadvantages of the different types of funding systems that are used around the world.

Figure 11: Relative performance orientation in the university funding mechanisms



Source: This figure is based on Jongbloed and Vossensteyn (2001) but with adjustments in Australia and New Zealand to reflect recent changes.

Table 3: Advantages and Disadvantages of Alternative Approaches to University Research Funding

	Advantages	Disadvantages
Performance-based	- Meritocratic - Improve individual and institutional performance - Competition may lead to increased efficiency - Encourages research to be properly finished and disseminated - Public accountability for government funds - Encourages explicit research strategies - Provides mechanism to link research to government policy - Concentration of resources enables best departments to compete with world leaders.	- High cost - May lead to homogenization of research - May discourage more risky research - May encourage 'publicaton inflation' - May encourage 'academic' research - Separates research from teaching - Rewards past performance - May lead to excessive government 'interference'
Educational size	- Low cost of administering - Allows universities to invest new people/research areas - Allows for long-term research - Encourages diversity in research - Allows any university to do research - Can encourage integration of research and teaching - Protects autonomy	- Little direct incentive to improve research performance - Excessive power to those who distribute core funding - Little accountability may lead to 'ivory tower' research - Reinforce stereotype of lazy academics - May be no relation between students numbers and research effort - Distribution of resources unrelated to government policy - Spread resources too thinly implies inability to compete with world-leading institutions

Source: Adapted from Geuna and Martin (2003)

b) Other International Differences

There exist other methods of improving research output. Other institutional changes, such as the degree of centralization of universities or the extent to which market forces are allowed, could influence how much research is undertaken. Another institutional feature that could, for example, explain Sweden's high research output is that researchers do not teach undergraduates[27] and therefore would likely have more time to devote to research.

In many countries research is undertaken in separate research institutes that are not necessarily affiliated with universities, such as in France and Germany. In Germany much research is undertaken in the various Max Planck Gesellschaften, and in France research is undertaken in the laboratories of the CNRS and INSERM. This phenomenon is reflected in Figure 12, which shows the split between R&D performed by governments themselves and R&D performed in higher education. France and Germany undertake a large proportion of their government-funded research outside of universities, and impact heavily on the EU average.

The reliance on research institutes has several implications:

a) There may be less competitive pressure to produce research output. The institutional structure may therefore explain the relatively poor performance of Germany and France shown in Figure 2;

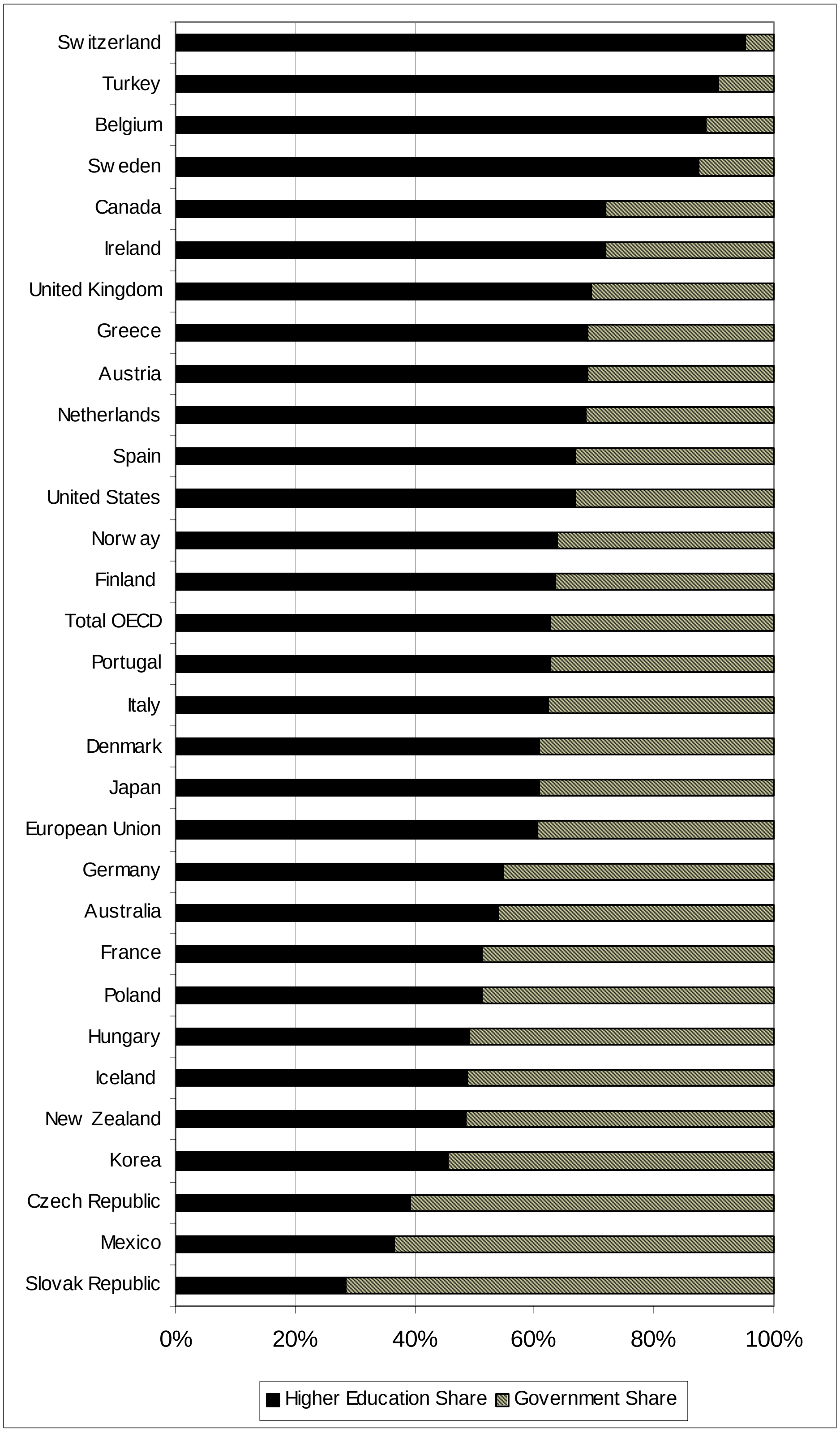
b) It is noticeable that there is a reasonable correlation between the relative importance of universities in government expenditure on research (from Figure 12) and the ranking for citations (from Figure 4). It should be emphasized however that publications are not the be-all and end-all of research systems;

c) The need for evaluation of universities' research output will vary across countries depending on the prevalence of research undertaken in universities. There is less need to evaluate university outputs if research is not undertaken there;

d) Output of research institutions directly controlled by government may be influenced by wider social goals, e.g. environmental protection or health and safety issues. Some cross-country regressions suggest that expenditure on government R&D has contributed to productivity growth, but the contribution has been greater in those countries where university research constitutes a greater share of public-sector research.[28]

[27] Mentioned in Henrekson and Rosenberg (2000). There also seems to be a concern in Sweden that recent reforms aimed at increasing undergraduate enrollment may be damaging Sweden's research environment.
[28] Guellec and de la Potterie (2001). Their research also suggests that public research needs to be mediated through private research: there are complementarities between private and public R&D.

Figure 12: Share of expenditures by performing sector: higher education and government



Source: OECD(2003)

5. Evaluation from the University's Perspective[29]

From the point of view of society at large, increased evaluation of research outputs could be a way of increasing output. But there are advantages to evaluation from the universities point of view as well.

Evaluation within a field of study or a university is well accepted. Its purpose is to establish the legitimacy of a research work for an internal audience, i.e. to ensure that there is quality improvement. There is however an increasingly important need to establish the legitimacy of research for an external audience—quality assurance—which is necessarily more difficult to do as the external audience has less expertise.[30] External legitimacy is required because of the increasing amount of resources required by researchers, resources that raise questions about the opportunity cost of such research. Furthermore, such a public affirmation of research quality would attract high-quality staff and students.

An essential challenge is how to ensure that legitimacy through external accountability can reinforce internal accountability. In the U.S., for example, accreditation by an independent education board for teaching quality may increase internal accountability for increased teaching standards.[31]

Internal – quality improvement

The first stage of internal accountability within a university evaluates how researchers carry out their 'mission'. Accountability is legal and financial: how are resources used and are they used for their intended purpose?

Academic accountability is explaining what has been done with those resources: has research quality gone up, for example. This second type of accountability is obviously more difficult because it refers to outcomes. Evaluation through peer review tries to establish quality through accountability to peers.

It is argued that a set of social norms exists within each scientific community, anchored by peer review and the referee system, which assure the quality of research output. This closed quality evaluation system is and was key to establishing the authority of science. Hence the essential, and idealized, feature of science has been its autonomy.

But can these internal processes, which are obviously necessary, convince external audiences?

External-quality assurance

Although the imperative to study science may be obvious to many, the need for resources to be expended in its pursuit suggest that the merits of research need to be justified in a concrete fashion because alternative, and more tangible, objectives also require public expenditure. Is pursuing science and spending money on it worthwhile?

---

[29] This section draws on Trow (1996) and Hansson (2002).

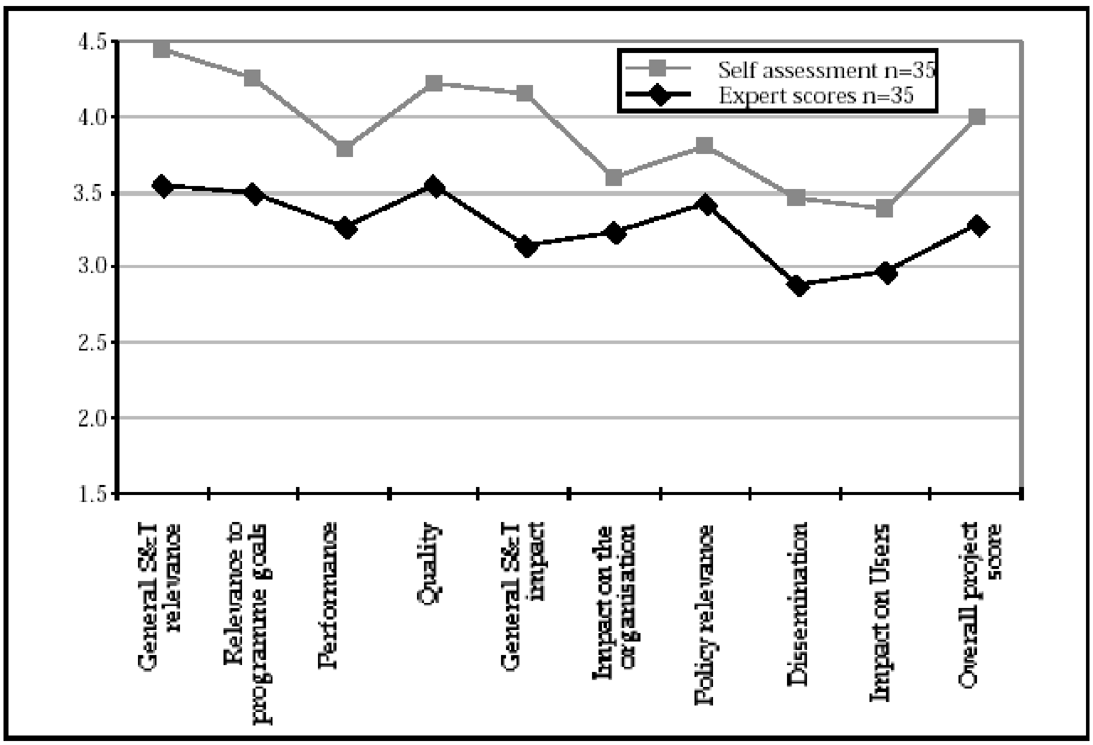
[30] These two roles of quality assurance and quality improvement (note that these are not mutually exclusive) were used within the Irish assessment system. See Michael Laver, "The Assessment System in Ireland", pp 19-21 in DISRRP (2003).

[31] There is also an increased amount of performance-based funding by the U.S. states although it is not clear what are the indicators upon which performance is based. See Burke and Minassians (2003).

Legitimizing research is important in mobilizing external resources. Resources enable researchers to produce results, which in turn are demonstrated in scientific publications. Publication allows the researcher to make a credible claim on future resources. This credibility arises because of the act of publication. But publication is an act internal to a discipline. Only experts within that discipline—peers—are able to evaluate the paper. So a problem arises because this process of evaluation through peer review is necessary to establish the quality of papers within a discipline but it cannot fully justify the output to outsiders.

Some of the reason for skepticism with respect to self-assessment is shown in Figure 13. This figure compares evaluation of a Swedish program done by those internal and external to the program. Those doing the internal assessment are uniformly more optimistic than external evaluators.

Figure 13: Experts' scores compared to respondents' scores for project performance in the Swedish Energy and Environment Program



Source: pp. 77 in Arnold and Guy (1997).

It is therefore in research institutions' own interest that they assure their funders of the importance of their contribution. There are essentially three fundamental ways in which universities are linked to their surroundings and their 'supporting societies':

a) Accountability
   Accountability is the obligation to report to others, particularly governments.

b) Trust

Trust is required when universities receive resources in return for producing services (research and teaching) without the need to account in detail for the use of those funds, e.g. when universities receive block grants.

c) Markets
Universities can interact through markets by, for example, charging tuition fees and teaching students in return or undertaking contract research.

Accountability, markets and trust can be substitutes or they can be complements: ‘trust but verify’ as President Reagan used to say. Around the world the prevalence of these interactions between universities and society differ. *Prima facie*, the market is the predominant link in the U.S. But there is also active accountability by universities to parents and particularly to alumni groups to reinforce the trust that these outsiders have in the institution. Because of the intensity of market forces in the U.S., the universities are keen to explain to their constituents—to be accountable—what they are doing with their resources in order to build up trust.

The U.K. university system operates in a world, according to political scientists, where the government has withdrawn its trust but has also limited the role for the market (at least with respect to tuition fees). Consequently formal accountability is substituted.

In Canada at the moment it would seem that the scientific enterprise relies largely on trust. But it would seem that the weakest of these science-society links to guarantee is trust. It is not wholly within anyone’s power to ensure that outsiders can trust scientific output and trust, if it can be established, can be fragile. As mentioned in the introduction, the extreme skepticism currently in Europe over genetically modified food may stem from past scientific failures, such as the recent scare over mad-cow disease in the U.K.[32] Because of the fragility of trust, sociologists believe there is a global change from relationships based on trust to a system based on collaborative assurance.[33] Increased accountability reinforces trust.

In continental Europe, accountability is concentrated through control of funding. However, the continental approach lead to difficulties in justifying research. As Campbell and Federer note,

> Among policy makers—and experts in general—there is some dissatisfaction with the structures and performance of Germany’s contemporary university system. And this dissatisfaction to explains a large extent also the reluctance of the public to increase the public funding base for university activities, such as university research. The traditional resistance of parts of Germany’s academic university community against systematic evaluations ... has had the negative effect—as some experts would pit it—that a ‘distrust gap’ between the public and the universities could not be bridged. … one could cite Wilhelm Krull who asserts that certain analysts are willing to phrase the West German mass university as the Achilles’ heel of Germany’s national R&D system.[34]

Evaluation can be an important force in building trust.

---

[32] For some discussion of the scientific problems in assessing the risk, see presentation of Millstone (2003) and Vogel (2001).
[33] Guston (2000).
[34] pp. 46 Campbell and Felderer (1997).

6. Evaluation Methods

This section addresses three questions:

1. What output is to be evaluated?
2. Who is to be evaluated?
3. How is evaluation done?

What is to be evaluated?

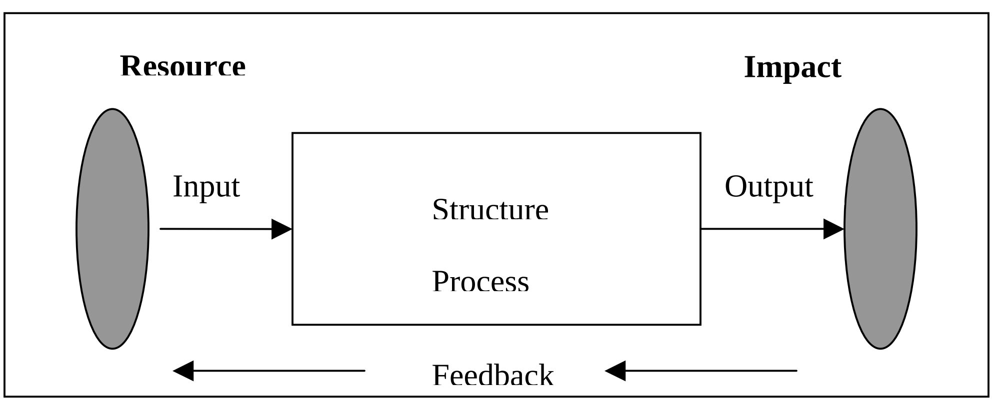
The essential first purpose of evaluation is to address organizational issues in academia.[35] In particular, evaluations

- Optimize the organization of institutions
- Reinforce principles of accountability
- Emphasize the application of explicit and rational criteria for decision-making and policy-making

The objective of evaluation is not a single flawless study, but "a succession of studies that critique and improve upon each other as they collectively advance toward norms of formal evaluation methodology."[36] There are advantages and drawbacks of all of the measures and methodologies. Given these pros and cons, many of the methods are complementary. Potential weaknesses of numerical indicators could be remedied by reviewing them by peers, for example.

Before getting into detail on what is being evaluated, it may be useful to present a stylized model on how the research process can be represented as a system. In Figure 14, resources—dollars and time—are inputs to be used within an organization in producing outputs and these outputs, e.g. an academic paper,[37] in turn have impacts on society, e.g. energy-saving technology. The ability of the institution to transform inputs into outputs depends on the structures and processes in the institution but the entire transformative process also depends on the workings of a state's economy.[38]

Figure 14: A simple systems model and variables of interest



Source: pp. 74 in Arnold and Guy (1997).

[35] This discussion draws on David F.J. Campbell (2002).
[36] Amy Glasmeier, Mel Mark and Irwin Feller, quoted in Feller (2000).
[37] Trained students are of course also an important output.
[38] Clearly the feedback of impacts on resources is important in this system. The interactions within the system are far more complicated than depicted here, interactions that make innovation a recursive dynamic system rather than a simple linear 'science supply-push' system.

The key role of valuation is to assess whether or not goals are being met. In order to fulfill this objective, it obviously follows that the desired goals have to be clearly defined. At this stage, we will merely observe that the government desires beneficial impacts on society.

Figure 14 can also be used to illustrate some of the concepts used and to demonstrate certain difficulties in understanding the operation of the system. The traditional approach to examining the efficacy of a research system is to quantify the inputs used by institutions, such as the number of researchers or dollars expended. Clearly, these measures do not necessarily illuminate how the research system works or whether there is a beneficial impact.

Figure 14 also illustrates timing issues. In general earmarked funding is transferred prior to research taking place. Hence it is always allocated on an *ex ante* basis but not necessarily on an *ex post* basis. However, the effectiveness of these funds transferred to universities can only be evaluated *ex post*. To a certain extent, *ex post* evaluations are used to evaluate the effectiveness of *ex ante* methods of allocating research resources.

Various types of measures may be associated with different parts and participants in the system. Interest in different types of information or data will differ according to the interest of the observer. A typology of data might consist of the following:

- *Management statistics*: data that describes the various characteristics of an institution, e.g. number of researchers, number of graduate students;
- *Management information*: when statistics are linked together principally for internal and external reporting and monitoring purposes but could be used to inform management decisions;
- *Performance indicators*: a quantitative or qualitative information that provides strategic information about the functioning and effectiveness of a system or organization

Clearly, the government, as a representative of society, has an interest in performance indicators so that it may monitor the well-functioning of the system as a whole. The challenge for evaluators therefore is to obtain measures that serve to inform on whether pre-determined strategic and social goals are being met. An example provided by AUCC of a system introduced in Canada is shown in Appendix 1.

A further distinction in evaluation is whether it performs a *summative* or a *formative* function. Summative evaluation involves assessing the performance of a unit (department, university, etc.) based on accumulated evidence in comparison with that of similar units. The goal of this type of evaluation is to find out what is going on. Formative evaluation on the other hand aims to assist the unit in meeting its goals, in obtaining data that could improve performance. To a certain extent summative evaluation can be an ongoing process to monitor a system whereas formative evaluation can be more exceptional. During the 1980s, for example, countries as diverse as the U.S. and Sweden questioned the validity of technology funding, which led to increased summative evaluation—what are we getting for our money?

For government, consequently, evaluation of university research could imply <u>measurement</u> of its

- Quality
- Efficiency
- Relevance

- o Viability
- o Effectiveness

A few or all of these could in principle be assessed. The U.K.'s Research Assessment Exercise (RAE) explicitly measures only quality whereas the Netherlands measures the first four.[39] A comprehensive list of objectives is given in Table 4.

Table 4: Typical evaluation issues

Issue	Question
Appropriateness	Was it the right thing to do?
Economy	Has it worked out cheaper than we expected?
Effectiveness	Has it lived up to expectations?
Efficiency	What's the return on investment? (ROI)
Efficacy	How does the ROI compare with expectations?
Process Efficiency	Is it working well?
Quality	How good are the outputs?
Impact	What has happened as a result of it?
Additionality	What has happened over and above what would have happened anyway?
Displacement	What hasn't happened which would have happened in its absence?
Process Improvement	How can we do it better?
Strategy	What should we do next?

Source: pp. 72 in Arnold and Guy (1997).

Who should be evaluated?

A fundamental difficulty with evaluating the research output of a university is that a) there is fundamental heterogeneity between disciplines (e.g. cost and time to publication); and b) universities will have different specializations. Combining these two factors implies that research output will differ across universities not only because of the differences in the quantity and quality of output but also because of composition differences within universities.

Table 5 shows how citation rates differ across subjects: there is a much greater rate of citing papers in molecular biology and genetics than in mathematics, for example. Different citation rates could be caused by fewer researchers in a field, greater difficulty in making breakthroughs, different rates of complementarities between research within a field, being more mature subject areas or a host of other explanatory factors. These difficulties are replicated when evaluating the research output of a university as a whole as opposed to an individual department.

Consequently producing indicators for a university as a whole would seem to be counter-productive. Such an indicator would likely be prejudiced against higher-cost disciplines where there are long lags to publications. Indicators for universities as a whole are not informative because they would not give an opportunity for less successful departments to learn. Each discipline should be evaluated separately. But one should emphasize that in evaluating each discipline, it is not necessarily only the academics within each discipline that are being assessed.

Table 5: Average citation rates for papers published 1993-2001 (across all countries)

[39] For definitions in the latest Dutch process, see VSNU/ KNAW/NAW (2003).

Fields	All Years	Fields	All Years
All Fields	8.31	Mathematics	2.47
Agricultural Sciences	4.31	Microbiology	13.38
Biology & Biochemistry	14.86	Molecular Biology & Genetics	23.99
Chemistry	7.46	Multidisciplinary	3.48
Clinical Medicine	9.82	Neuroscience & Behavior	15.7
Computer Science	2.27	Pharmacology & Toxicology	8.82
Economics & Business	3.82	Physics	6.81
Engineering	2.84	Plant & Animal Science	5.72
Environment/Ecology	6.98	Psychiatry/Psychology	7.61
Geosciences	7.35	Social Sciences, general	3.14
Immunology	18.92	Space Science	10.88
Materials Science	3.77		

Source: (ISI Citation Database)

The implications of research evaluation are relevant to individual university departments, to universities, to granting councils and to governments. These institutions' choices—in how much money they give and in how they allocate it—are important in determining outcomes. The object is to evaluate the outcomes of choices made by those who receive financing. Hence it is appropriate to evaluate an individual researcher's research output if it is that researcher that has received a grant, as is currently done by the granting councils. If a university as a whole receives funding, e.g. through a block grant, then it is legitimate to evaluate use of the money by that university as a whole. By looking at the performance of a university in individual subjects one is able to ask whether a university allocates money to its most productive use.[40] Do the choices made by granting councils increase research excellence? This type of question can only be answered by looking at individual subjects.

How to evaluate research output?

How research is evaluated depends to a large extent on what one wants to evaluate. Recalling the terminology of Figure 14, Figure 15 shows the type of research evaluation that is appropriate. On the horizontal axis, the evaluator may have an interest in inputs, outputs or impacts. Government should be interested in the last two.

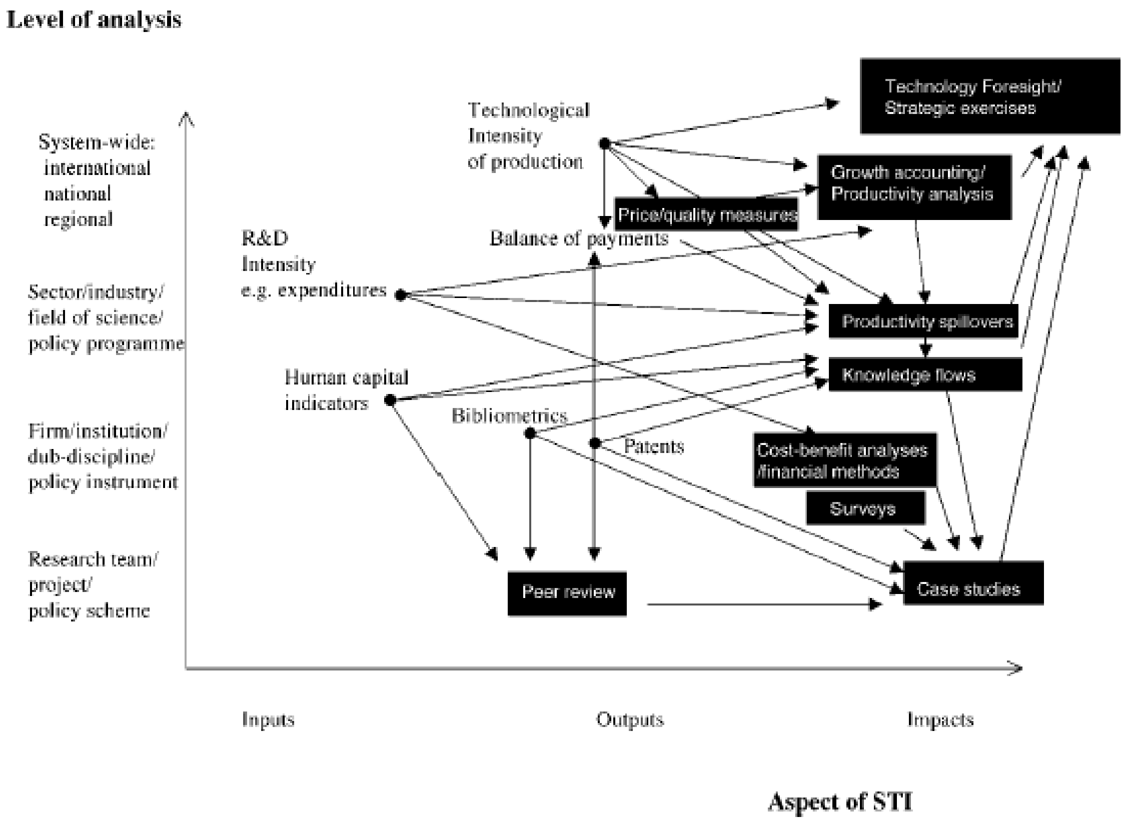
On the vertical axis is the degree of detail in which one is interested, from the performance of an individual researcher to the performance of a country's research system. Consequently, if the evaluator is interested in the output of a researcher then peer review is the most appropriate but if the evaluator is interested in the impact of the researcher's output on society then a case study may be more appropriate. Note again that these methods are not mutually exclusive.

A government, in principle, should be interested in the impact of the research system as a whole, i.e. in the top-right part of Figure 15. However, clearly the different types of evaluations involve significant trade-offs. The amount of detail and the cost vary across each method. It is not practical to regularly redo growth accounting exercises but patent and bibliometrics data are

[40] For global rankings by subject, see http://adminsrv3.admin.ch/cest_ccs/cgi-bin/content.cgi?1bb207f3684b6e1ebbfc609396949dc69d8fc4c1d4c18fd7d7d78fc3c5d3d48fc6d28fd3c3c9c5ced4cfcdc5d4d2c9c58fc3c8c1cdd0c9cfced3bfccc5c1c7d5c58ec8d4cd9d969491969496959591969495949394949194 9d9892959099929297969195999493969491959dc58fc8c1cdd3d4c5d28fccc9d3d4d38fd5cec9d6bfd2c5c6 .

relatively easy to obtain. Consequently, even if the government is interested in the top-right hand corner of Figure 15, information from methods given in the rest of table may be sufficient to assure society that the research system functions well. The hope is that policies that encourage research output will have a positive impact on society.

Figure 15: Indicators and Evaluation Techniques for Research



Source: Kane (2001)

Definitions

Obtaining satisfactory economy-wide measures of research impact is likely to be costly and rife with uncertainty. Consequently, this section will start to explore measures of research output. Overall methods of evaluating outputs will be discussed before delving into more details in subsequent sections.

There are two ‘ideal type’ methods of evaluating university research output: *peer review* and *indicators*. There is no obviously correct approach and both these methods have been combined in the modified peer review method discussed below. Traditional peer review is a method by which a panel (one ore more) of peers—established experts—within a discipline evaluates scientific work. Indicators introduce a numerical estimate of the quality of research. In practice the indicator method is also a peer review process albeit indirect: the indicators are measures of success at the peer review method (i.e. number of publications) or in the influence papers have had upon the scientific community (i.e. citations).

An important distinction, obviously, is that the set of peers in a peer review evaluation is much smaller than the set of researchers who cite important papers. In this sense, citation impact may be more valid than a peer review process that only uses peers from within a small country (which is why some Scandinavian countries use an international group of peers).

The positive and negative implications of peer review and indicators can offset each other, as seen in Table 6.[41]

Table 6: Strength and weaknesses of methods of evaluating university research

	Numerical Indicators	Modified peer review
Advantages:	They are relatively objective and quite cheap to implement. Many countries around the world produce these types of statistics, e.g. it is the current approach in Australia.	It can integrate complexity.
Disadvantages:	They may be seen as superficial and therefore unable to fully represent the complexity of the research process. As such, academics may view them as lacking in legitimacy.	It is costly and potentially subjective.

Peer review, because informed participants undertake it, can integrate the complexity of the research process by, for example, evaluating the difficulty of tasks undertaken. However, the negative aspect is that even well informed participants can impart subjective notions into their evaluation, e.g. maintaining current research objectives rather than experimenting with untested methods.

Indicators, by their nature, are much simpler and therefore cannot replicate, for example, the degree of difficulty of the research undertaken. However, indicators do measure precisely and objectively what they are supposed to measure, even if there is debate about the implications of that measure.

Because of arguments given later on in this section the traditional peer review method is not usually used for evaluations of systems at a more aggregate level than the individual or small team of researchers. Table 7 gives the two common ways used to evaluate output. The U.K. and Hong Kong use a modified peer review method in helping to allocate funds. Modified peer review is a method that a) involves more than just peers and b) may look at wider issues than the pure quality of an academic paper. The Netherlands uses a similar method although it does not currently have any bearing on funding decisions. Some other countries use numerical indicators to allocate funds.

Table 7: Classification of universities according to evaluation methodology

Modified peer review		Numerical Indicators	
Used to allocate funds	Used to inform opinion	Used to allocate funds	Used to inform opinion
U.K. Hong Kong	Netherlands	Australia Poland	Finland Switzerland

[41] But both approaches can be combined. The U.K.'s RAE uses peer review to come up with an indicator of quality. The Dutch system uses bibliometrics to inform its peer review process.

			Denmark Japan

Within the broader context of research other evaluations methods could also be used, such as customer satisfaction data and economic impact studies. For instance, the most commonly used system in Finland to assess R&D projects is *users' evaluation*. Economic impact studies try to obtain some measure of the net present value, benefit-cost ratio or social rate-of-return of research projects. These types of evaluation may be easier to implement for a single institution rather than being developed to judge an entire system however.

Different types of evaluations will now be discussed in some more details. Only those techniques that the author thinks could be developed fairly easily for all university departments are considered. Consequently user surveys and more complicated cost/benefit analysis will not be considered here.

Indicators

As mentioned above, the role of indicators is to supply data that informs users on whether strategic goals are being met. The challenge therefore is to ensure that data provided would actually give the intended information. An example of how indicators could leas users astray is the use of class size as a proxy for the quality of education, as the empirical links between class size and education quality are pretty weak.[42]

The following section concentrates on publication output. Obviously publications may not be a sufficient indicator but one could argue that increased publication is a necessary goal of academic research. There may be other aspects of academic output that could be used as indicators. Obvious candidates would be the number of masters and doctoral students, and the time that it takes for them to complete their studies. Other indicators that could be easily measured are indicators of commercialization, such as patents and license revenue.

One should also note that although there may be problems with one indicator system relative to another, it may be more informative to look at the behaviour of an indicator over time.

Publication[43]

Publication counts on their own are obviously too simplistic: an article in the *Journal of Snow Economics* cannot be as important as an article in *Econometrica*. Consequently, one option is to weight publications according to the journal in which they appeared. The mean citation rate of all articles contained in the journal could be used, which is published annually in the Science Citation Index Journal Citation Report.

The principal problems with an approach based purely on publication are that:

- It leaves out other measures, e.g. patents;
- It may be affected by biases in the publication process;
- It may have difficulties with joint-authorship;
- Publication rates may differ with department size.

[42] AUCC (1995).
[43] Pp. 104 in CPB

Citation Analysis[44]

These indicators mostly use the Science Citation Index from the ISI.[45]

Problems with this approach may be that:
- It is biased towards publications in the English language;
- Cannot distinguish between positive and negative citations;
- Limitations of the database.

On the positive side:

1) Bibliometrics may not be suitable for evaluating an individual's output because it is difficult to take quality into account. But if publications are aggregated for a discipline and a normal distribution is assumed for quality then bibliometrics can serve as a useful source of information at an aggregate level.[46]

2) Given that peer-reviewed publications form the basis of the analysis, bibliometrics is, more or less, a *quantified* quality system, or an indirect peer review method.

3) To a certain extent, the number of papers that can cite an important paper is much greater than the number of peers involved in an assessment for a granting council. Further an international audience could cite such a paper. In principle, therefore, wide citation of a paper may be a greater sign of esteem than the opinion of two peers on a panel.

Benchmarking

In the U.S. the National Academies' Committee on Science Engineering and Public Policy (COSEPUP) tried to establish whether the U.S. was a leader in various scientific fields. The objective was to answer the question of how the federal government could "gauge the overall health of research enterprise as a whole and in its parts and determine whether national funding is adequate and whether it supports national research objectives"[47].

To do this COSEPUP established panel of peer reviewers. The panels consisted of researchers in the field, in fields close to that subject and research users. Foreign scientists were also included.[48] These panels were then asked about the position about the current and expected position of U.S. research. To get to their judgements, the panels used:
- The virtual congress

  All participants were to identify the best-of-the-best in each field by their own judgement and by polling other leaders in the field. The idea was to identify who were the individuals that would participate in a session that would consist of only the top 5-20 individuals in the world.
- Citation analysis
- Journal publication analysis
- Quantitative data analysis (e.g. funding)
- Prize analysis

[44] pp. 105 in CPB
[45] For a discussion of best approaches to undertake citation analysis, see Kostoff (2002).
[46] Campbell and Felderer (1997).
[47] Statement of Fox (2000).
[48] In the mathematics panel, three of the twelve were foreign.

- International congress speakers

The Swiss, using a soccer analogy, benchmark their universities against a Champion's League of universities. The Swiss recognize that although they do well in country-to-country comparisons (i.e. in comparing averages), they also have to compare their institutions to the best in the world (i.e. compare the whole distribution).[49]

'Traditional' Peer Review

Note that peer reviews can use quantitative information about publications and citations, a process that is usually called *informed peer review*, and discussed in the next section.

Figure 16 gives some of the positive and negative arguments for peer review viewed from the academic community. The debate centers on the degree of subjectivity in the process.[50] Do the small group of peers act as impartial and independent observers. This subjectivity can be mitigated by having international experts on the review committees, as is done in Sweden. However, one of the complaints in Sweden is that all projects still tend to obtain a score of 4/5.[51]

One difficulty that may need to be emphasized is the often-heard complaint that the peer review process can become conservative.[52] Reviewers are often well established in their field and may therefore be more inclined to a particular view of which research fields are fruitful. A complaint that then follows from this observation is that the peer review process is biased against interdisciplinary research, which is an increasing feature of the research environ. There may be self-censorship of radical ideas

Figure 16: Positive and negative arguments on peer review

Positive arguments for peer review	Negative arguments for peer review
1) An effective resource allocation mechanism	Partiality of peers to impact the outcome for non-technical reasons
2) An efficient resource allocator	An 'old boy' network to protect established fields
3) A promoter of science accountability	A 'halo' effect for higher likelihood of funding for more visible scientists/departments/institutions
4) A mechanism for policymakers to direct scientific effort	Reviewers differ in criteria to assess and interpret
5) A rational process	The process assumes agreement about what good research is, and what are promising opportunities
6) A fair process	
7) A valid and reliable measure of	

[49] "Même si l'on estime que la place scientifique Suisse n'est pas (encore) en péril, il faut se regarder de toute complaisance à l'égard de nous-mêmes. On se demandera plutôt si les comparaisons entre pays, qui sont généralement favorables à la Suisse, reflètent suffisamment l'évolution de choses. Pour répondre à cette question, il faudrait une observation plus rapprochée, englobant la comparaison avec les régions scientifiques et les institutions de recherche les plus performantes du monde." Da Pozzo (2000).
[50] For some of the literature on peer review, see http://www7.nationalacademies.org/core/peer_review_background.html .
[51] DEST, Chapt 3.
[52] Some of the discussion here is based on Hansson (2000).

scientific performance

Source: Kostoff (2002)

A difficulty with peer review, which may not be perceived by academics, is that peer review is an internal process. As such, peer review may be useful in determining who obtains research projects since informed judgment is required to identify potentially fruitful topics. The negative implication of an internal system is that such qualitative evaluation is not very persuasive to an external audience.[53] Indeed, in other countries the demand for evaluation has come from those outside the research community.

A second difficulty with peer review is cost. The amount of time necessary to undertake a thorough review of the work to be evaluated may be large. Using outside peers to review the functioning of a department may also be large.

Bearing in mind these pros and cons, there is some consensus in the literature that peer review is well adapted to[54]

1. Fields with a defined and delimited community of researchers with stability regarding journals and an articulated reward structure;
2. Fields in which problems of research are chosen by the scientific community itself;
3. Fields that can expand through more financing.

There is doubt that pure peer review is useful for processes that have wider objectives, such as evaluation of a research system or the allocation of funds across departments. Peer review cannot determine priority between basic and applied research, and rank topics of research across panel areas.

Kostoff (2002) concludes that the demand for a 'perfect' peer review is very costly to meet. The social consequences of the organization of the process are unknown but are probably not exactly productive because of the inherent conservative nature of the review process.

[53] Doremus (2002).
[54] Summarized in Sandstöm (2002).

Table 8: Strengths and Weaknesses of Various Evaluation Tools

Tool		Strengths and Weaknesses
Case Studies	+	Help understanding of complex processes
	+	Explore situations where interesting variables not predefined
	+	Can be structured and subjected to analysis
	+	Provide 'how to' understanding, for example of success-factors
	-	Highly dependent on evaluators' skills and experience
	-	Expensive if done in large numbers
	-	Hard to incorporate into routine monitoring
	-	Generate limited quantitative information
Bibliometrics	+	Useful for understanding some macro-level patterns, especially in more basic types of research
	-	Unhelpful where populations are small, as in a single research programme
	-	Focus on academic research. Under-values other types
	-	Differences in propensity to publish among disciplines and at different stages of the 'innovation cycle'
	-	Number of publications not a reliable indicator for quality or scientific progress
	-	Meaning of citation counts ambiguous: Positive or negative citation?
	-	Increasing share of collaborative, multi-authored publications in journals is hard to evaluate
	-	Bias towards English-language journals
	-	Under-represents new and emerging fields
	-	Gratuitous co-authoring and co-citation as authors manipulate counts. 'What you measure is what you get'
Co-word Analysis	+	Useful for mapping fields
	-	Immature technique. Meaning unclear for evaluation purposes
Peer Review	+	Informed, 'rounded' judgement, especially of scientific quality
	+	Can be systematised, checked and analysed to increase confidence in results
	-	Typically over-influenced by scientific quality criteria
	-	Qualitative, judgement basis leaves it open to criticism
	-	Problems of criterion-referencing and differing cultural behaviours
	-	'Group think' and social dominance effects within panels
	-	Risk of 'prisoner's dilemma' behaviour by peers
	-	Hard to apply to commercially-sensitive work
Patents Analysis	+	Useful for understanding some macro patterns and problems relating to programme appropriateness
	-	Inherently unlikely to captures outputs from basic research
	-	Patents indicate neither technical nor economic success
	-	Variations in national patent systems
	-	Variations in patenting propensities between countries, branches of industry and individual companies/institutions
	-	Tell nothing about non-patented or non-patentable aspects

Table 9: Strengths and Weaknesses of Various Evaluation Tools, continued.

Tool		Strengths and Weaknesses
User Surveys	+	Can provide a nuanced, quantified understanding of programme
	+	Collects direct process experience as well as indicators
	+	Can test and generalise case study and other findings
	+	Enables estimation and description of key impacts
	+	Provides quality control of programme management
	-	Subject to positive bias, reflecting users' appreciation of receiving resources and optimism about impacts
Cost/Benefit Analysis etc	+	Requires a rigorous and systematic analysis of costs and effects
	+	Forces the construction of a model of the action being evaluated
	-	It is possible to generate coherent but widely differing benefit/ cost ratios
	-	Focus on money creates a risk of under-valuing non-monetary costs and benefits
	-	Typically does not handle externalities/spill-overs
	-	Costly to collect the needed detailed benefit and cost data
	-	Difficult to deal adequately with multiple causality
	-	Deals poorly with multi-step causality

Source: Arnold and Katalin Balázs (1998)

'Modified' or 'Informed' Peer Review

Increasingly countries have been introducing a modified peer review process. Modified peer review is changed by

a) Gathering a different group of peers, and

b) Different aspects of the research system are evaluated.

This new process therefore takes place within a broader context and with more information. For example, the peers may evaluate research in the context of wider societal goals. They may also use bibliometrics in arriving at their conclusion.

Modified peer review can, for example,

- Involve not only scientific personnel. Industry or those applying research outcomes can be involved in evaluation, i.e. the 'users' are involved. This process has been undertaken in Sweden, the Netherlands and by SSHRCC in Canada. Representatives of other disciplines could also be included.

- The RAE in the U.K. is essentially a modified peer review where the objective is to arrive at an assessment of research quality. These peers can also use bibliometrics work to evaluate quality.

In other words the definition of peer or the definition of task have been significantly modified. This process seems to be used extensively in Scandinavia since the 1980s. When the pressure for evaluation grew in Denmark, as discussed in the next section, modified peer review was advanced as a compromise with the RAE developed in the U.K. This Danish model seems similar to that undertaken in the U.K. but the evaluators and specific procedures are open for negotiation.[55]

Comparison of Methods

To a certain extent the same complaints could be leveled against both the indicator and peer review systems. They both could have a bias towards research undertaken in the English language, against younger researchers and against more innovative but risky research. Roberto Perotti notes that output-based evaluation is based on noisy signals of quality, but that this must still be better than no signal at all. He notes the example of the RAE in the U.K. that “by trial and error, has worked remarkably well by almost unanimous consent”.

Researchers in the U.K. have been able to compare the results of the RAE, which is based on peer review, and those of a bibliometrics study.[56] They concluded,

- RAE grades and bibliometrics impact correlate well for the natural science
- There is almost no correlation in the arts and humanities
- Over the time period analyzed (1988 to 1996), there was increased publication and citation rates in the U.K. for the natural sciences
- The social sciences increased publication rates but have likely maintained their citation rates over the same period.

Using bibliometric methods is likely to be considerably less costly than using modified peer review, as in the U.K. However, given a fairly prevalent suspicion of bibliometrics as a robust signal of quality, it would seem appropriate that bibliometric analysis be conducted by specialized panels who could apply some degree of informed opinion to ensure that the numerical indicators are sound, as in the Dutch system.

Should resources be allocated according to measures of research output?

This subsection discusses whether resources should be allocated according to any measures of research output that would be developed.[57] Allocating resources according to measured output would introduce ‘high-powered’ incentives to increase those outputs that are measured. This incentive system has two implications:

1) Measured outputs differ, to some extent, from the underlying desired output. Measures are an imperfect proxy, e.g. publication as a measure of research activity. Consequently, high-powered incentives would increase measured output rather than necessarily increasing desired output. Hence, the desirability of high-powered incentives depends on the extent to which measured output reflects the desired output.

[55] Hansson (2000).
[56] Adams et al. (1997).
[57] Theoretical advances in incentives in principal-agent games with multitasking started with Holmstöm and Milgrom (1991).

2) Allocating effort to measured outputs may be to the possible detriment of those outputs or activities that are not measured. For example, if publications were measured but citations were not measured then there would probably be an incentive to increase the number of publications without worrying too much about quality.

   Such a change may be positive however if the current 'high-powered' incentives were directed at activities other than research. For example, if there were many incentives to increase undergraduate teaching (as is currently the case in some Canadian provinces) then, at the margin, research activity may suffer because the inadequacy of incentives. Introducing powerful incentives would remedy this effect.

It may be too crude to rely on numerical indicators as a measure of research output to justify initially introducing high-powered incentives. Clearly if a limited amount of additional funds were made available then such resources could be allocated by measures of research outputs. But the U.K. system allocates large quantities of resources according to research output and therefore establishes high-powered incentives. The U.K. approach can be justified since they established an extensive and expensive system of measuring research quality, which provides greater assurance that the measured research quality closely matches actual research quality.

This conclusion holds for measures of research activity such as scientific advance, but may be different for other forms of research 'output'. Important outputs of research are research students (masters and PhD) that spread the latest scientific advance. If the government were reasonably confident that increased postgraduate education were desirable in and of itself then high-powered incentives could be advanced in this area, i.e. by allocating resources based on the number of graduates as is done in Finland, Denmark, Australia, etc.

7. What has been the Impact of the RAE in the U.K.?

The U.K. introduced a system to evaluate research output in the mid-80s. The inception of the RAE was greeted with some skepticism. Furthermore it is likely that the initial attempts at evaluation were not particularly successful at encouraging increases in research quality. But now, despite not spending as much resources as many other countries, the U.K. produces high quality academic research, as seen in Figure 17. Furthermore the RAE does not seem to penalize multidisciplinary research, as seen in Figure 18.

Figure 17: What is the State of Research in the U.K.?

- The U.K. ranks thirteenth out of seventeen comparator group countries for expenditure on R&D across the economy
- Its share of PhD awards is similar to that of other countries, after taking population differences into account
- Its share of publications was second but has now dropped to third behind Japan
- The U.K.'s citation count is the second highest in the world at around 11% of world citations
- The U.K. is second in almost all fields except for mathematics (3rd) and Physical Sciences and Engineering (4th)
- Its global share of citations is increasing except for social sciences and engineering
- Business R&D is generally thought not to compare well internationally, except in the pharmaceutical industry.

Source: OST (2003).

Evaluations of the RAE have concluded that the institutional response to the RAE has been to:[58]

1. Make structural changes:
   a. Give research a higher priority;
   b. Develop internal process for evaluation;
   c. Selectively allocate resource for research;
   d. Place senior managers in charge of monitoring and managing the work.
2. Allocate RAE-generated research funds and other moneys to achieve the highest possible rating in subsequent evaluations;
3. Some gaming.

For individuals, these surveys conclude that academics in more highly rated departments are under pressure to publish in high-quality journals. Individuals in lower-rated department should publish more, regardless of location. Academics spend more time on research and research output has improved both quantitatively and qualitatively.

Looking at the research output of economics departments, Moore et al. (2002) suggest that those economists in highly ranked department significantly increased their publications in quality journals between 1992-1996 relative to 1980-1989. Academics in lower ranked department increased publications. Increased publication was a result of increased effort, not from the influx

[58] See reports cited in Moore et al. (2002)

of new higher-quality academics. Most of the change in behaviour was for those whose publication performance was below par prior to the evaluations. There was not much change for those individuals who were already producing high-quality publications. The authors note that the 1992 and 1996 exercises caused a cumulative productivity shock over an individual's career: cumulative research output in quality journals is higher than it would otherwise have been. The effect of the 1992 RAE caused per capita publications in quality journals to increase by approximately one publication during the four-year period following that exercise.

Figure 18: Interdisciplinary research in the RAE

There has been discussion of whether the RAE is biased against interdisciplinary research. This is a particularly important question with the increased amount of interdisciplinary work being undertaken, as mentioned in the introduction. The body running the RAE was concerned about this issue and commissioned an independent report. The conclusion of this report was:

- Interdisciplinary work is indeed important: around 80 per cent of research are engaged in this type of research
- There is a belief that the RAE inhibits such research
- But there is no evidence to support this claim for the 1996 RAE.
- There were some procedural difficulties for referring research to a particular subject panel for evaluation.
- But these problems were rectified in 2001 by
    - Changing the composition of review panels to include wider areas of research,
    - Change in methodologies,
    - More reporting of interdisciplinary research

In the abstract, one could argue that interdisciplinary research would not be penalized by a well-designed system. The wider group of peers in a modified peer review should mitigate a bias against subject specificity, particularly in comparison to pure peer review. Secondly, if there were complementarities across subjects then one would expect the performance assessment to increase as a consequence. It would be the responsibility of the university to internalize these spillovers.

The RAE undertaken in 2001 showed a dramatic increase in the (self-assessed) quality of research. A full 40 per cent of departments reviewed were in the top two rankings. A House of Commons committee undertook a review of these results to see whether they were genuine.[59]

The Committee concluded that

- There was some 'gamesmanship' undertaken to boost ranking. However the RAE did reflect a genuine improvement in quality of research.
- Benefits had been obtained, principally through improved research management and targeting areas of excellence.
- Costs included the administrative burden, distortions in research objectives and the implications for those academics and departments that did not obtain high rankings.

[59] See Hagan (2002).

The Commons committee concluded that additional resources were required to support university research in light of the improved performance.

An independent report in the U.K. came to the following conclusions:[60]

- The international comparative performance of the U.K.'s higher education research is 'extremely competitive'. Performance has 'measurably and progressively improved over the last 15 years.
- U.K. research provides tangible value for money in terms of quantity as well as excellence and there has been a measurable gain in research performance.
- In case studies, the researchers found that university staff spoke of a change in research culture since 1986 because of the institutional structures and procedures that were introduced to guide research.

Costs included

- Diversion of some resources away from capital projects and towards increasing research performance because of insufficient funding,
- Increased workload.

Others have noted the administrative burden that assessments placed on universities.

A summary of the costs and benefits given by a U.K. academic is given in Table 10. It would seem that some of the evidence presented above suggests that the advantages of the RAE have led to tangible improvements in research. On the negative side, some of the drawbacks listed in Table 10 may be related to design issues rather than to the underlying ideas of evaluation. For example, if there were sufficient time between evaluations then there would be no bias against safer research. Furthermore, there is no convincing evidence of complementarities between teaching and research, much as one may wish them to exist.

Table 10: Advantages and Drawbacks of the RAE System of Research Funding Allocation

Advantages	Disadvantages
Rewards good research performance – seen as meritocratic. As such, the allocation can be seen as 'fair'.	Can encourage game playing, i.e. undertaking ancillary effort that do not result in improved research performance
Encourages creation of larges research groups	May lead to excessive concentration of research – less diversity in research
Provides incentives to improve individuals research performance	May lead to safer research
Competitive system – encourages completion of research, weeds out poor quality, better research management	Encourages strong researchers to move to highly rated institutions, but this may weaken synergies between teaching and research
Can link government policy priorites and research funding streams	Can lead to excessive government involvement.
	Rewards past performance
	May not encourage new entry into research.
	Cost

Source: Hare (2002).

In conclusion, from the various discussions of the U.K. system, there seems to be a consensus that the RAE has improved research outputs. This benefit has come at a certain cost. The direct financial costs do not seem to be onerous: around 1 per cent of the total research budget.

[60] Evidence (2002)

However the opportunity cost of the time academics have to allocate for the administrative burden of the RAE may be more significant.

In an editorial in *Science*, Jonathan Adams (2002) concluded,

> U.K. research improved relative to world benchmark, reversing declines in the early 1980s. Average research performance, on the basis of publication and citation data, gained progressively from 1986 to 2000. … The RAE created incentives that channeled natural research competitiveness into a pervasive driver of excellence. It overcame initial opposition in the community by showing that those who deliver good research have little to fear and much to gain from well-structured scrutiny.

The U.K. has a gold-plated system. With comprehensive evaluation and large amounts of resources being allocated by this evaluation, there are high-powered incentives to increase research.

Such an incentive structure may not be appropriate in all circumstances. Consequently other countries that followed the U.K.'s lead initially adopted methods that either had less powerful incentives or were aimed at achieving lower cost. Because of the modified peer review system's cost, Australia's version of the RAE initially relied on publication counts. Nevertheless there has been considerable evolution in other countries evaluation systems, as the next two sections will discuss.

8. Evolution of evaluation systems

This section is intended to give some sense of the dynamics that could be in play if an evaluation system were introduced, based on the experience of some other countries. Details of evaluation systems in other countries are given in Appendix 3.

Australia

Australia introduced a system of evaluating research output in the early 1990s. At the beginning, a simple count of publications was used as an indicator.[61] The result of this policy can be seen using bibliometrics: Australia's share of publications in major journals has increased from 2.2 per cent to 2.8 per cent over the 1990s.[62] Australia represents around 1.2 per cent of global GDP. However, as might be expected from using a simple publication count, quality seems to have declined, as measured by citation measures.

This research system is now being reviewed.[63] Although many have noted the cost of introducing a scheme similar to the U.K.'s RAE, the Australian National University (for example) advocated the introduction of modified peer review rather than continuing with indicators because such a system:[64]

- Provides a mechanism that realistically reflects research quality, rather than relying on quantitative surrogates, in the allocation of research funds. Discipline committees in an RAE system exist specifically for this purpose.
- Government Higher Education research funds would be more likely to go to universities undertaking the highest quality research.
- Compliance by universities with an expanded and more inclusive list of research performance indicators would provide clearer research management information for internal management of university research efforts, as well as evidence for external reporting and communications.
- Performance against such indicators would also provide strategic marketing material for universities.
- Excellent research centres in smaller universities are more likely to be recognized fiscally for their research performance.
- Provides a mechanism for the assessment of research performance by experts in relevant fields.

Denmark

Evaluation is not a static process that is imposed and then maintains that form *ad infinitum*. In Denmark, the initial impetus for evaluation came from politicians in the early 80s.[65] This first stage of introducing evaluation was a top-down approach of government mandating evaluation. The government introduced modified peer review in 1983, following the Swedish model, which involved a wider population than just scientists. Funds are allocated between research centres

[61] Data on the number of publications and resources expended by university are available at http://www.dest.gov.au/highered/ki_reforms/allocations_2003.xls.
[62] Report available at http://www.science.org.au/butler/.
[63] http://www.dest.gov.au/highered/ki_reforms/default.htm .
[64] http://www.dest.gov.au/highered/ki_reforms/submissions/pdf/k83.pdf.
[65] Evanthia Kalpazidou Schmidt, 2003, "Evaluation and Science Policy", pp. 7-36 in DISRRP (2003).

through modified peer review.[66] Danish universities are now attaching new external users and employers to the advisory boards for research and education.[67]

Performance indicators were introduced in 1981 but were opposed by the Ministry of Education until 1989 when the ministry also adopted them. By the mid-90s however, research bodies themselves were initiating evaluations. Evaluations are undertaken on a subject basis, e.g. social science research in 1997. Evaluation was a strong legitimation device since newer institutions introduced many of the evaluations. Universities also increasingly initiated evaluation. Each department now prepares an annual performance review.

In 2000 there was a move to 'development contracts': agreements between each university and the Ministry for Research. These contracts focus on success criteria and measurement of outputs. A proportion of research funds is now given according to amount of external funding.

However, universities and education bureaucrats did remain intransigent and a proposal to move to the U.K. system in 1995 was rejected. But the system seems to be open to change: "As the universities open up to the world and enter into new partnerships and ways of interacting with external parties, it must be expected that evaluation of research quality and productivity will take on new forms that to a higher degree mirror legitimate external needs and requirements to the research. … As an example of increased openness and transparency for the wider public and stakeholders, the universities have created public research databases where citizens can search information about university activities, either to find relevant partners from e.g. university departments, or to get research results on concrete themes."[68]

A taximeter, which is based exclusively on outputs (number of students graduated), allocates teaching grants.

Netherlands

A system of evaluating education was established in 1988 for evaluating research in 1993. Once every three years research units in the Netherlands produce a self-evaluation. Its purpose alternates between preparations for an external evaluation and serving as an internal mid-term evaluation (i.e. each university has an external evaluation every six years). External evaluation is undertaken by international committees of peers, which uses bibliometrics used as well.[69]

The evaluation system aims at three objectives with regard to research and research management:

- Improvement of the quality of research through an assessment carried out according to international standards of quality and relevance;
- Improvement of research management and leadership;
- Accountability to higher levels of the research organizations and funding agencies, government, and society at large.[70]

---

[66] DISRRP, 2003,
[67] pp. 95 *supra*. For reforms of the granting council system in Denmark, http://www.videnskabsministeriet.dk/fsk/div/forsksoejlen/OECD5.pdf .
[68] pp. 95, Government of Denmark (2004).
[69] For details, see Roel D. Bennink, "Quality Assessment of Research in the Netherlands", pp. 28-33 in DISRRP.
[70] http://www.knaw.nl/cfdata/publicaties/detail.cfm?boeken__ordernr=90000091.

The impact of assessments in the Netherlands has been to:[71]

- Increase cooperation between researchers, as research programs not researchers are evaluated;
- Increase the publication rate, especially in international journals with high impact factor;
- More power to university administrators (at the expense of department chairs). Assessments provide a solid basis for managers to use as instruments of quality control. There is also a shift in power to management of faculty and the university;
- Importance of *research policy* is enhanced;
- Greater reputation to those who obtain better assessments. The prestige of researchers has increased;
- Public reports have "made it virtually impossible for weak and unproductive groups to continue to exist unnoticed"[72].

But there may have been:

- Greater research orthodoxy
- Greater attention to research in English
- Loss in the unity between teaching and research

Focus has been on management of research programs. Use for accountability and allocation is limited.

Switzerland

A federal act, past in 1999, noted that "a more specific definition of the organizational structures tends at output-oriented subsidies and an efficient evaluation-system".[73] The Center for Science and Technology Studies (CEST) evaluates research subsidized by the federal government, including all of the government's own research and particular scientific areas. In order to guarantee its independence, CEST is funded with the Federal Interior Department and the Federal Economy Department.[74] CEST's goals include:

- An empirically sound comparative analysis and assessment of the strengths and weaknesses, potentials and prospects of Switzerland as [a] place of scientific research and technological development;
- The evaluation of both effectiveness and efficiency throughout the Swiss science system as far as federal funding is involved;
- The development of an evaluation culture within the Swiss science system.[75]

CEST undertakes evaluation of universities and/or departments based on mutual consent.[76]

For an example of the output from the CEST comparing universities, see Appendix 4.

---

[71] Monique Leijennar, "Who Wants Research Assessments and What For?", pp. 8-11 in DISRRP (2003) and Roel D. Bennink, "Quality Assessment of Research in the Netherlands", pp. 28-33 in DISRRP (2003).
[72] Bennink, *supra*.
[73] Garke (2000).
[74] CEST (2003a).
[75] Website at http://www.cest.ch/en/das_cest/einheit_evaluation.htm.
[76] For details of the research performance of Switzerland, see CEST (2003b).

9. Conclusion

Canadian research seems to be quite good but there does not seem to be 'excellence' in research at the moment. However, this perception may result from the paucity of data. The available citation and publication data seem to suggest that the Canadian average is doing well compared to the average in other countries. But from looking at the ranking at the top end of the quality distribution, Canada does not seem to be faring quite as well.

Research output is a function of resources spent and the microeconomic incentive structure. Increased resources alone will not necessarily increase and improve academic output. An important tool being deployed world wide, and discussed in this paper, is the implementation and evolution of methods to evaluate research output. Many countries that perform well on measures of research output have methods to evaluate research output. Countries that introduced these systems have subsequently strengthened their systems by introducing improved incentives. Policymakers in these countries must believe there are tangible benefits from evaluating research output through enhanced incentives to produce high-quality research output.

Evaluations are important as incentives but at a more fundamental level they provide data on the research activity within a country. If there is no transparent and objective way of examining research activity, it is difficult to determine whether the research system is 'working' and where and how it can be improved.[77]

[77] Some performance indicators are being introduced in Canada: in Ontario these performance indicators are based on undergraduate teaching outcomes: http://osap.gov.on.ca/eng/not_secure/urlsuniv.htm; in Alberta research performance is measured as well: http://www.uquebec.ca/conf-quebec/actes/s12.pdf. Data on the University of Alberta is available at http://www.president.ualberta.ca//pdfs/FavoriteChartsJune2004WordVersion.pdf for example. Data on the commercialization of research is not available at a detailed level in Canada. A report to evaluate technology transfer in Canada notes: "There is a need for more publicly available, comprehensive information on the performance of Canadian universities in technology transfer, upon which informed decisions on public policy can be made. Although the AUTM Surveys captured most of the universities that attract most (approx. 90%) of the research funding in Canada ... they comprised only about one-half of the universities in the country…Statistics Canada surveys appear to gather information from most of the universities in Canada, but their reports do not make public any information on individual institutions … only aggregated data, which do not reveal any information about the tremendous variation that we have shown exists among the universities." Clayman (2004).

References

AAS, 2003, "Policy Statement on Research and Innovation in Australia", *Australian Academy of Science*, September, http://www.science.org.au/academy/media/10september03.pdf.

Jonathan Adams, 2002, "Research Assessment in the U.K.", *Science* 296, pp. 805.

Jonathan Adams, Rim Bailey, Louise Jackson, Peter Scott, David Pendlebury and Henry Small, 1997, "Benchmarking of the International Standing of Research in England", Centre for Policy Studies in Education, University of Leeds and Institute for Scientific Information, Philadelphia.

Erik Arnold and Ken Guy, 1997, "Technology Diffusion Programmes and the Challenge for Evaluation", http://www.oecd.org/dataoecd/34/40/1907926.pdf.

Erik Arnold and Katalin Balázs, 1998, "Methods in the Evaluation of Publicly Funded Basic Research: A Review for OECD", Brighton: Technopolis, http://www.technopolis.co.uk/downloads/eval_bas.pdf.

Erik Arnold, Stefan Kuhlman and Barend van der Meulen, 2001, "A Singular Council: Evaluation of the Research Council of Norway", Technopolis, http://odin.dep.no/archive/kufvedlegg/01/03/243AR028.pdf.

AUCC, 1995, "A primer on performance indicators", *Research File* 1, 2, http://www.aucc.ca/_pdf/english/publications/researchfile/1995-96/vol1n2_e.pdf.

BMBF, 2002 , "Report of the Federal Government on Research 2002" pp. 27ff, http://www.bmbf.de/pub/bufo2000_englisch.pdf.

BMBF, 2003, "Facts & Figures Research 2002" pp. 210 ff. http://www.bmbf.de/pub/facts_and_figures_research_2002.pdf.

Joseph C. Burke and Henrik Minassians, 2003, "Performance Reporting: 'Real' Accountability or Accountability 'Lite', Seventh Annual Survey 2003", The Nelson A. Rockefeller Institute of Government, http://rockinst.org/publications/higher_ed/7thSurveyfinal.pdf.

David F.J. Campbell, 2002, "Conceptual Framework for the Evaluation of University Research in Europe", http://www.gwu.edu/~cistp/PAGES/campbell_2002.pdf, CISTP Working Paper.

David F.J. Campbell and Bernhard Felderer, 1997, "Evaluating Academic Research in Germany: Patterns and Policies", Institute for Advanced Studies, Vienna working paper no. 48.

CEST, 2003a, *Rapport annuel 2002*, Bern : Centre d'études de la science et de la technologie.

CEST, 2003b, *Place scientifique Suisse, 2001, Développements de la recherche en comparaison internationale sur la base d'indicateurs bibliométriques 1981-2001*, Berne, CH : Centre de compétence en scientométrie de CEST, http://www.cest.ch/Publikationen/2003/CEST_2003_2.pdf.

CHEPS, 2003, *Higher Education Policy Issues and Trends, An Update on Higher Education Policy Issues in 11 Western Countries, 2003*, Enschede, Netherlands: Center for Higher Education Policy Studies, http://www.utwente.nl/cheps/documenten/engreport03monitorupdatereport.pdf

Bruce P. Clayman, 2004, "Technology Transfer at Canadian Universities: Fiscal Year 2002 Update: A Report for the Canada Foundation for Innovation", available at http://www.sfu.ca/vpresearch/vprreports.htm.

CNER, 2002, *Évaluation De La Recherche Publique dan Les Établissements Publics Français,* Comité National D'Évaluation de La Recherche, http://www.cner.gouv.fr/fr/pdf/bib.pdf.

CNER, 2003, *La Recherche Publique Française: Une Évaluation, Rapport au Président de la République*, Paris, France : Comité National d'Évaluation de la Recherche, http://www.cner.gouv.fr/fr/pdf/rpr.pdf.

CSST, 2003, "Formation et Recherche en Suisse : Les priorités du CSST", Conseil Suisse de la Science et de la Technologie, http://www.swtr.ch/swtr_fr/pdf/swtr_prioritaeten_f.pdf .

François Da Pozzo, 2000, "Benchmarking international dans le secteur de la recherche : Les institutions suisses et la 'Ligue des Champions'" La Vie économique, CEST, No. 9/2000.

DISRRP, 2003, *The Use of Evaluations in Europe: Report from the European RTD Evaluation Network Meeting*, Aarhus: Danish Institute for Studies in Research and Research Policy, http://www.afsk.au.dk/ftp/Evaluation/2003_2.pdf.

Paul Doremus, 2002, "Federal Agency Research Assessment: NIST", G-8 Working Group on Research Assessment.

EC, 2002, *Towards a European Research Area, Science, Technology and Innovation, Key Figures, 2002*, Brussels: European Commission, http://europa.eu.int/comm/research/era/pdf/benchmarking2002_en.pdf.

Evidence, 2002, "Maintaining research excellence and volume: A report by Evidence Ltd. To the Higher Education Funding Councils for England, Scotland and Wales and to Universities UK", http://www.hefce.ac.uk/Pubs/rdreports/2002/rd08_02/rd08_02.pdf. Full report at http://www.evidenceuk.com/downloads/maint-research-exellence.pdf.

Irwin Feller, 2000, "The Academic Policy Analyst as Reporter: The Who, What, and How of Evaluating Science and Technology Programs", Proceedings from US-EU Workshop Learning from Science and Technology Policy Evaluation, Bad Herrenalb, Germany, September, http://www.cherry.iac.gatech.edu/e-value/bh-proceed/02-Frameworks.pdf.

Forfas, 2002, "Baseline Assessment of the Public research System in Ireland in the areas of Biotechnology and Information and Communication Technologies", http://www.forfas.ie/publications/0304_publicresearchsys/BaselineAsses.pdf.

Marye Anne Fox, 2000, “Experiments in International Benchmarking of U.S. Research Fields”, The Committee on Science Subcommittee on Basic Research, United States House of Representatives, October 4, 2000, http://chancellor.ncsu.edu/speeches/houseofrepoct00.htm.

Esther Garke, 2000, “Swiss Higher Education: A Survey”, http://www.bbw.admin.ch/html/pages/services/publikationen/uni/higher_education/text-education.html.

Aldo Geuna and Ben R. Martin, 2003, “University Research Evaluation and Funding: An International Comparison”, *Minerva*. 41, 4, http://www.sussex.ac.uk/spru/publications/imprint/sewps/sewp71/sewp71.pdf

Michael Gibbons, Camille Limoges, Helga Nowotny, Simon Schwartzman, Peter Scott and Martin Trow, 1994, *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*, London: Sage Publications.

Anthony Giddens, 2003, *Runaway World: How Globalization is Reshaping our Lives*, 2nd Ed., New York: Routledge.

David H. Guston, 2000, “Retiring the Social Contract for Science”, *Issues in Science and Technology*, http://www.nap.edu/issues/16.4/p_guston.htm.

Government of Denmark, 2003, *Promoting university interaction with business and community - a comparative study of Finland, Sweden and UK*, Ministry of Science Technology and Innovation, 2003, Government of Denmark, http://www.videnskabsministeriet.dk/cgi-bin/doc-show.cgi?doc_id=177151&leftmenu=PUBLIKATIONER.

Government of Denmark, 2004, “Danish universities in transition - Background reports to the OECD examiners panel 2003”, Ministry of Science Technology and Innovation, Government of Denmark http://www.videnskabsministeriet.dk/fsk/div/forsksoejlen/OECD8.pdf . (Entire report available at http://www.videnskabsministeriet.dk/cgi-bin/doc-show.cgi?doc_id=190990&leftmenu=PUBLIKATIONER).

Government of Japan, 2001, “The Science and Technology Basic Plan (2001-2005)”, Cabinet Office, http://www8.cao.go.jp/cstp/english/basicplan01-05.pdf.

Government of Japan, 2003, *The Final Report on "A New Image of National University Corporations"(A report from "Study Team Concerning the Transformation of National Universities into Independent Administrative Corporations")*, Ministry of Education, Culture, Sports, Science and Technology, http://www.mext.go.jp/english/news/2003/07/03120301/004.htm. Details on evaluation available at http://www.mext.go.jp/english/news/2003/07/03120301/004/005.pdf .

Pat Hagan, 2002, “Bargain-basement research”, *The Scientist* April 26.

Finn Hansson, 2000, “Research evaluation: Controlling or developing the production of knowledge?” The peer review system in the changing knowledge production system”, paper presented at Evaluation 2000, AEA-conference, Hawaii, http://www.remap.dk/departments/mpp/remap/publications/aeapaper.doc .

Finn Hansson, 2002, “Best practice in research evaluation? How to evaluate and select new scientific knowledge by introducing the social dimension in the evaluation of research quality”;

Paul G. Hare, 2002, “The U.K.’s Research Assessment Exercise Impact on Institutions, Departments, Individuals”, paper presented at OECD Conference on Higher Education, 2002.

Sven Hemlin and Sören Barlebo Wenneberg, 2002, “From Quality Control to Quality Monitoring and Organisational Learning, Copenhagen Business School working paper 13/2002.

Adam B. Jaffe, Manuel Trajtenberg and Rebecca Henderson, 1998, “Universities as a Source of Commercial Technology: A Detailed Analaysis of University Patenting, 1965-1988”, *Review of Economics and Statistics* 80, 119-127

Magnus Henrekson and Nathan Rosenberg, 2000, “incentives for Academic Entrepreneurship and Economic Performance: Sweden and the United States”, Institute for Industrial Economics working paper, http://swopec.hhs.se/iuiwop/papers/iuiwop0530.pdf.

Diana Hicks, Tony Breitzman, Dominic Olivastro and Kimberly Hamilton, 2001, “The Changing Composition of Innovative Activity in the U.S. – A Portrait Based on Patent Analysis”, *Research Policy* 30, 4, 681-703.

Bengt Holmstöm and Paul Milgrom, 1991, “Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design”, *Journal of Law, Economics and Organization* 7 (special issue), 24-52.

ICSTI/Forfas, 2002, “Measuring and Evaluating Research, Report of the ICSTI Task Force”, Irish Council for Science, Technology and Innovation, http://www.forfas.ie/icsti/statements/stiresearch02/020827_icsti_measuring_and_evaluating_research.pdf.

Dominique Guellec and Bruno van Pottelsberghe de la Potteire, 2001, “R&D and Productivity Growth: Panel Data Analysis of 16 OECD Countries” *OECD Economic Studies* 33, 103-126, http://www.oecd.org/dataoecd/26/32/1958639.pdf.

Ben Jongbloed and Hans Vossensteyn, 2001, “Keeping up Performance: An International Survey of Performance-Based Funding in Higher Education, *Journal of Higher Education Policy and Management* 23, 2127-145.

Aidan Kane, 2001, “Indicators and Evaluation for Science, Technology and Innovation”, *mimeo*, http://www.economics.nuigalway.ie/downloads/kane/kaneicsti02.pdf.

Ronald N. Kostoff, 2002, “Citation Analysis of Research Performer Quality”, *Scientometrics* 53, 1, 49-71, available at http://www.onr.navy.mil/sci%5Ftech/special/354/technowatch/docs/citation_analysis_paper2.doc

Wilhelm Krull, 1999, “Facilitating Change in the German Research System: Capabilities, Collaboration and Competition”, paper presented on “Evaluation of Science and Technology in the New Europe", Berlin, http://www.isi.fhg.de/ti/Ealuation.pdf. See also http://www.mpiwg-berlin.mpg.de/ringberg/sessionIV.html.

Terttu Lukkonen, 1997, "The Increasing Professionalisation of the Evaluation of Mission Oriented Research in Finland: Implications for the Evaluation Process", http://www.oecd.org/dataoecd/4/5/1823310.pdf.

Craig McConnell and Robert H. March, 2001, "Bringing Reasons and Context to the Science Wars", *Physics Today*.

Erik Millstone, 2003, "BSE: Risk, Science and Governance", presentation at SPRU, http://www.sussex.ac.uk/spru/documents/031123.ppt.

William J. Moore, Robert J. Newman, Peter J. Sloane and Jeremy D. Steely, 2002, "Productivity Effects of research Assessment Exercises", Department of Economics, Louisiana State University Working Papers 2002-15, http://www.bus.lsu.edu/economics/papers/pap02_15.pdf.

Helga Nowotny, Peter Scott and Michael Gibbons, 2001, *Re-thinking Science: Knowledge and the Public in an Age of Uncertainty* Cambridge, U.K.: Polity Press.

NSF, 2002, *Science and Engineering Indicators, 2002*, National Science Foundation http://www.nsf.gov/sbe/srs/seind02/c5/fig05-34.htm

OECD, 1999, *University Research in Transition*, Paris: OECD.

OECD, 2001a, "OECD Science, technology and Industry Scoreboard, 2001, Towards a Knowledge-Based Economy", http://www1.oecd.org/publications/e-book/92-2003-04-1-7294/GA-13-1.htm.

OECD, 2001b, MSTI database, May, http://www1.oecd.org/publications/e-book/92-2001-04-1-2987/gA-9-2-a.htm

OECD, 2003, MSTI database, May, http://www1.oecd.org/publications/e-book/92-2003-04-1-7294/Annex_tables_excel/At5.1_e.xls

OMB, 2002, *The President's Management Agenda*, Office of Management and Budget, http://www.whitehouse.gov/omb/budget/fy2002/mgmt.pdf.

OST, 2003, "PSA Target metrics for the U.K. Research Base", Office of Science and Technology, Department of Trade and Industry.

Parliament of Canada, 2002, *Canada's Innovation Strategy: Peer Review and the Allocation of Federal Research Funds, Report of the Standing Committee on Industry, Science and Technology*, Ottawa: Government of Canada, http://www.parl.gc.ca/InfoComDoc/37/1/INST/Studies/Reports/INSTRP10-e.htm.

A. Abigail Payne and Aloysius Siow, 2003, "Does Federal research Funding Increase university Research Output?", *Advances in Economic Analysis and Policy* 3, 1.

Roberto Perotti, 2002, “The Italian University System: Rules vs. Incentives”, *mimeo*, European University Institute, http://www.iue.it/Personal/Perotti/papers/Universita_6_4-02.pdf.

Dino Rizzi and Paolo Silvestri, 2002, “The Evaluation of the Italian University System: A Recent History”, Centro di Analisi delle Politiche Pubbliche, http://www.capp.unimo.it/pubblicazioni/materiali_di_discussione/RizziSilvestriEnglishVersion.pdf .

Ulf Sandstöm, 2002, “Peer Review – A Process for Credibility and Quality”, Swedish Institute for Studies in Educaion and Research, http://www.forsk.dk/dkeuformand/sandstroem.pdf.

Luis Sanz-Menéndez, Cecilia Cabello and Fina Antón, 1998, “Technology foresight in its science and technology policy context in Spain”, CSIC Institute for Advanced Social Studies, Madrid, http://www.iesam.csic.es/doctrab1/dt-9808.htm.

Luis Sanz-Menéndez, 1995, “Research Actors and the State: Research Evaluation and Evaluation of Science and Technology Policies in Spain”, *Research Evaluation* 5, 1, 79-88, available at http://www.iesam.csic.es/doctrab1/dt-9512.pdf.

Uwe Schimank and Markus Winnes, 1999, “National Report: Federal Republic of Germany, European Comparison of Public Research Systems” Max PlanckInstitute for the Study of Societies. http://www.mpi-fg-koeln.mpg.de/berichte/NationalReport.pdf.

Alan Sokal and Jean Bricmont, 1998, *Fashionable Nonsense: Postmodern Intellectuals’ Abuse of Science*, New York: Picador Press.

STPCF, 2000, *Review 200: the Challenge of Knowledge and Know-how* Helsinki: Science and Technology Policy Council of Finland, http://www.minedu.fi/tiede_ja_teknologianeuvosto/eng/publications/Review_2000.html#4.

STPCF, 2003, *Knowledge, Innovation and Internationalisation* Helsinki: Science and Technology Policy Council of Finland, http://www.minedu.fi/tiede_ja_teknologianeuvosto/eng/publications/review_2003.pdf.

Martin Trow, 1996, “Trust, Markets and Accountability in Higher Education: A Comparative Perspective”, *Higher Education Policy* 9, 4, 309-324.

Frank van der Most and Barend van der Meulen, 2001, “RCN’s role in the dynamics of research: a scientists’ perspective, Sub-report for the evaluation of the Norwegian Research Council,” University of Twente, http://odin.dep.no/archive/kufvedlegg/01/03/243BR032.pdf and http://odin.dep.no/ufd/norsk/forskning-old/tema/014021-990023/index-dok000-b-n-a.html.

David Vogel, 2001, *The New Politics of Risk Regulation in Europe*, London: London School of Economics. http://www.lse.ac.uk/collections/CARR/pdf/Disspaper3.pdf.

VSNU/ KNAW/NAW, 2003, “Standard Evaluation Protocol, 2002-2009, For Public Research Organizations”, available at http://www.knaw.nl/publicaties/pdf/90000091.pdf.

Appendix 1: Conceptual Representation of the Results-Based Management Structure (RBM)[78]

This appendix is comprised of two elements: a chart which explains key RBM terms and a Logical Framework Analysis (LFA) outline with instructions. All applications should be written using a RBM approach. At the Letter of Intent stage you will be asked to define your results in RBM terms. At the Full Proposal stage you will be asked to complete a full LFA for your project.

Inputs (Resources)
Inputs are the direct or indirect human, organizational or material resources required to implement and carry out the planned activities of the program or project.

Activities
Activities are the actions required to produce results.

Outputs
Outputs are the immediate measurable results necessary to produce the outcomes. Outputs are a direct consequence of the implemented activities.

Outcomes
Outcomes are specific measurable institutional or community level results that will be produced by the end date of the project, such as new programs or processes that will be sustained after the life of the project. Outcomes are the logical result of project outputs. Outcomes reflect the project purpose(s).

Impacts
Impacts are the long-term socio-economic results or changes that usually happen beyond the life of the project and are attributable to the achievement of outcomes. Impacts involve changes in living conditions of populations in developing countries. Impacts reflect the project goal.

□ □ (□)

Performance indicators should be identified for each output and outcome. In some cases it may also be appropriate to identify performance indicators for intended impacts. Indicators are proof or evidence that a result has been accomplished. They answer the question AWhat will be observed if the result is achieved?@ If a result is to improve or increase knowledge or capacity, ***baseline data*** may need to be collected to allow measurement of what has changed. Indicators should be simple and efficient, but they should clearly show whether or not the intended result was achieved.

[78] http://www.aucc.ca/upcd-pucd/_doc/applications/rbm-b-e.doc .

Appendix 2: Finland

Finland has made innovation a central part of its policy to improve standard living. As part of its concentration on innovation, evaluation of research is playing a key role, as discussed in Figure 19.

Figure 19: Evaluation in Finland[79]

Impact analysis and evaluations will continue

One important factor for the development of research is constant appraisal of performance and impact. There have been significant reviews of government policy, such as an impact analysis of the programme for additional research funding 1997–1999.

The state and performance of the Finnish research system as a whole was evaluated by the Academy of Finland in 1997, 2000, and 2003. The Academy has commissioned discipline assessments since the early 1980s; these have had a significant effect on the development of the disciplines.

The Academy of Finland will be evaluated by an international expert panel in 2003–2004.

Tekes has had external evaluations of all the national technology programmes. Tekes' own operations have been evaluated in several reviews conducted by external teams. The Ministry of Trade and Industry has published an evaluation of the Finnish innovation support system in 2003.

In spring 2001 the Ministry of Trade and Industry and Tekes launched a new research programme on interaction between technology, business and society.

The government research institutes were systematically evaluated during the latter half of the 1990s.

Performance evaluation is an integral part of university and polytechnic operations. Insitutional evaluations were conducted in all the universities in the 1990s, and many study fields were also evaluated. Evaluation played a particularly important role in the establishment of the polytechnic network in Finland during the nineties.

A Higher Education Evaluation Council was set up in 1996 to assist in the evaluation of higher education institutions. It has had a particular role in organising evaluations, making them a permanent element of higher education, and arranging follow-up to evaluations.

[79] From pp. 19, *Research in Finland*, http://www.research.fi/research_in_finland.pdf.

Appendix 3: Methods used by other countries

Table 11: State of play in other countries[80]

Name	State of Evaluation
Germany	No evaluation covering all universities. Some evaluation of individual universities and by individual Länder. A commissioned report recommended "new, more effective ways of monitoring and evaluating performance, including external participation."[81] There is evaluation of more scientifically oriented institutes, for example the Max Planck Gesellschaft (MP), commencing in 1997 and to be repeated every five years, focused on their scientific merits.[82] Note that much research in Germany takes place in research institutes.[83] An amendment in 1998 by the Bundestag allows for comprehensive evaluation of education and research.[84] 'Immunity' of professors to external evaluation is abolished. Starting in 2000 there were extensive system-wide and program evaluations undertaken.[85] However, "in spite of the described intense political regulation and control of universities, government actors by now have not gone very far in the direction of a regular and comprehensive evaluation of teaching and research performance. It is not to be expected that evaluation exercises like the ones that took place in Great Britain will be installed in Germany in the near future. There is nothing more than a variety of fragmented activities and approaches going on in the different states and universities. This means that information on relative performance as a crucial precondition for a competitive allocation of resources to universities, departments, or individual professors is relatively vague and incomplete."[86]
France	Resources for research are allocated by a) granting councils (e.g. CNRS) and by contracts between a central agency and the universities. The negotiations for these contracts are based on development plans and research evaluation. A report by the Comité National d'Évaluation de la Recherche concluded that there should be increased evaluation of research in the French system.[87]
Italy	Since 1998 evaluation is supposed to have taken place using measures indicating

[80] Data in this table comes from various sources such as Geuna and Martin (2003) and CHEPS (2003).
[81] Krull (1999). See also http://www.mpiwg-berlin.mpg.de/ringberg/sessionIV.html.
[82] An overall conclusion of international expert panels was that Germany has a good scientific position, but few centres of excellence. The entire system evaluation led to some reallocation of funding and a merger between institutes, which were incremental changes to the system rather than radical shifts in the PSRE landscape. http://trendchart.cordis.lu/Reports/html/4.2.html#germany.
[83] BMBF (2002).
[84] There was a constitutional ban on evaluating universities.
[85] BMBF (2003).
[86] pp. 186, Schimank and Winnes (1999).
[87] "Le CNER recommande que toute procédure d'évaluation et que les résultats auxquels elle conduit soient les plus transparents et explicites possibles, tan pour en assurer l'objectivité que pour permettre un éventuel débat évalué-évaluater et une meilleure prise en compte de résultats. … Le CNER recommande la mise en place d'une structure d'évaluation des équipes et laboratoires de recherche unique, extérieure aux organismes eux-mêmes, susceptible de faire appel autant que de besoin à des compétences externes (le plus souvent hors de frontières française) d'une part, aux responsables des établissements d'autre part.", pp. 157-159, CNER (2002).

	how much research support an academic had obtained. It is not clear if this evaluation has taken place.[88] There are also incentive problems in that external evaluations are actually done by units internal to the universities. [89] Perotti notes the "persistent dreams and obsessions typical of academic cultures that do not appreciate the value of competition and individual initiative in research" in Italy.
Ireland	The Universities Act, 1997, established mandatory quality assurance/quality improvement exercises. These were 'not modelled on the U.K. experience … but based on that used by most universities in Europe'. These reviews take place one subject at a time and each subject is reviewed every five years.[90] Informally the Irish departments were rated on the U.K.'s RAE scale, as many of the reviewers were from the U.K.[91] There are many one-off studies.[92] A study was conducted in 2002 that entailed bibliometrics analysis and panel reviews of the data. Subject areas were individually reviewed, particularly biotechnology and ICT. Review concluded that - There should be improved international benchmarking - A full-blown RAE could correct for problems identified with the evaluation methodology but this was expensive. Increased evaluation would take place as Ireland reorients towards science technology and innovation.[93]
Norway	There is a six-year cycle of evaluating institutions supported by the Research Council of Norway. There is also a rolling evaluation of disciplines.[94] These reviews have no impact on funding. An evaluation of the RCN concluded that "the council's overall spend on evaluation is low, compared with the informal OECD norm of 1-2% of research budgets, and appears overly focused on peer review." [95]
Finland	Some funds (3% of total) are allocated according to performance indicators (on quality and impact of research and teaching)[96] but there is no special evaluation for this process. Evaluation of particular research fields, one at a time, has taken place since 1983,

[88] Perotti (2002).
[89] Rizzi and Silvestri (2002).
[90] UCD News, 1998, "Quality Assurance – into the future for Ireland's universities". For more details, see Irish Universities Quality Board at http://www.iuqb.ie/.
[91] Michael Laver, "The Assessment System in Ireland", pp. 19-27 in DISRRP(2003).
[92] Forfas (2002).
[93] ICSTI/Forfas (2002).
[94] For an discussion of university system in Norway, including its low expenditure on research, see van der Most and van der Meulen (2001).
[95] pp. 87, Arnold, Kuhlman and van der Meulen (2001).
[96] More details on pp. 15 of Geuna and Martin (2003)

	which includes evaluation of university research. The University of Helsinki undertook a research assessment exercise in 1999.[97] All government research institutes were evaluated between 1995 and 1999. Evaluations are undertaken by a) external consultants and b) international peers.[98] There is an emphasis on openness.[99] A proposal to undertake a RAE as in the U.K. was shelved after political pressure. A report by the Science and Technology Policy Council of Finland (headed by the prime minister advocated, "[w]ithin sectoral research, the uncommitted research funds at the disposal of ministries must be evaluated - as the last significant part of the public research system still outside systematic evaluation."[100]
Austria	A Universities Act was passed in 2002. Traditionally universities' budget was based on negotiation between the university and the Ministry of Science. From 2005, 20 per cent of the budget will be based on performance indicators including research performance.[101]
Sweden	Up to the 90s the Swedish education seems research seems to have been left to researchers. During the 90s there was a concerted effort to promote interaction between universities and industry. Strategic research foundations were established in the 90s. Allocation between subjects was governed by permanent peer review panels chosen by academic communities.[102] These foundations placed great weight on evaluating outcomes of the programme with equal weight to scientific quality and to relevance. In 1997, a legal obligation was placed on universities to communicate their research results to the surrounding communities. In 2001, research councils were integrated in the Swedish Research Council. The Council notes, "One of the Swedish Research Council's functions is to analyse and evaluate Swedish research and research policy. The results become part of its documentation for decisions on various strategic issues. Activities include investigations and statistical surveys on research and its financing, and independent scientific evaluations of different research fields. The Council also advises the Government in matters relating to research policy."[103]
Spain	National Evaluation and Prospective Agency (ANEP) established in 1987 to encourage and undertake evaluations.[104] ANEP assists ex post evaluations when universities have volunteered to be assessed, e.g. Universidad Autónama de Barcelona.[105]
Portugal,	Under EU pressure because they are large recipients of structural funds, these

[97] http://savotta.helsinki.fi/halvi/keho/researce.nsf/6f5754c598d9f43785256207004f8789/$FILE/Research%20assessment.pdf

[98] Terttu Lukkonen (1997) and STPCF (2003).

[99] See Finnish Higher Education Evaluation Council at http://www.kka.fi/english/index.lasso?cont=english. For details of Finnish research, see http://www.research.fi/index_en.html.

[100] STPCF,(2000).

[101] pp. 17-22, CHEPS 2003

[102] http://www.vr.se/english/index.asp.

[103] http://www.vr.se/english/index.asp?id=572.

[104] For the history, see Sanz-Menéndez, et al. (1998).

[105] Sanz-Menéndez (1995).

Greece	countries have increasingly adopted research evaluation policies. In Portugal in general the government is said to have 'some impatience with the process and to consider more demanding processes of assessment…a view among certain public opinion quarters [is] that the process is too soft and without serious consequences for the institutions."[106]
Poland	Numerical indicators of research performance used to allocate funds. The process is still evolving.
Slovakia	Since 1992 there has been evaluation of research (based on indicators) and conducted every three years
New Zealand	20% of funds allocated through a contestable pool evaluated through peer review. Proportion may increase to 80% after review.
Japan	Japan passed a law in July 2003 to reform its universities. The reform is intended to allow mergers of universities in order to strengthen their position in global rankings based on assessments by third parties.[107] In particular, there will be comprehensive evaluation of education and research by independent third parties, and the results will be made public. "With regard to the basic role and nature of evaluation subjects, (1) Each university shall carry out rigorous self-examination and evaluation of the degree to which medium-term goals have been achieved, utilizing a variety of external evaluations and third-party evaluations, (2) The National Institution for Academic Degrees shall carry out evaluation from a specialist perspective concerning items relating mainly to education and research, based on the self-examinations and evaluations of each university, (3) The National University Evaluation Committee (tentative name) shall carry out a comprehensive evaluation on the administration of the national university corporation (tentative name) as a whole, respecting the results of the evaluation by the National Institution for Academic Degrees concerning education and research."[108] … "Each university shall utilize the evaluation results to contribute to improvements in education, research and other activities, and to examine its basic philosophy and long -term goals. Evaluation results shall be reflected in the content of subsequent medium-term goals and medium-term plans. When doing so, due consideration shall be given to ensure that their content is developmental and realistic, based on the evaluation results. Evaluation results shall be reflected in the calculation of operational grants for subsequent medium-term goal periods. When doing so, to promote the individualistic development of universities and foster a competitive environment, calculations must be carried out to appropriately reflect the results of evaluation of education, research and other activities."[109]

[106] CHEPS (2003).
[107] *Sydney Morning Herald*, December 10, 2003.
[108] Pp. 48, Government of Japan (2002).
[109] pp. 50, Government of Japan (2002)..

	Evaluation of ministries' research was first introduced in 1997 but is not budget to have had a sufficient impact. Consequently increased emphasis is going to be given to evaluation.[110]
Taiwan	Taiwan seems to be in the process of introducing a system similar to the U.K.'s RAE.
Hong Kong	Similar to RAE in UK. Undertaken since 1993. [111]

[110] Evaluation is extended to ministries and to major research projects, as outlined in pp. 29ff. Government of Japan (2001).
[111] Details available at http://www.ugc.edu.hk/english/documents/papers.html#papers.

Appendix 4: Switzerland

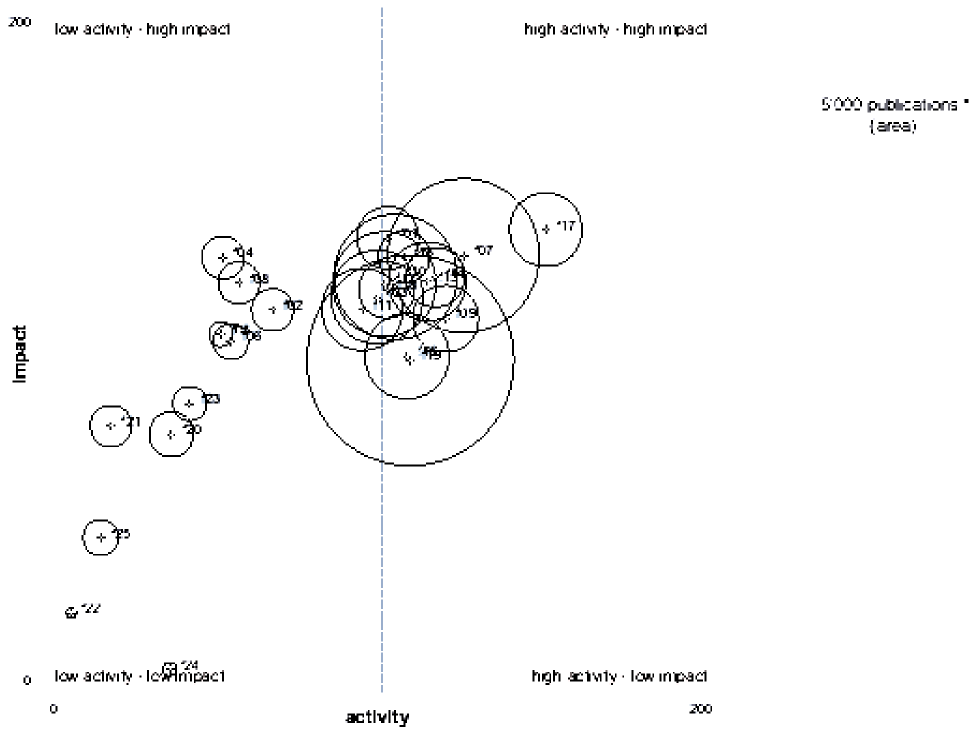
Figure 20 gives a picture of the research output of Swiss research in terms of citations. Each circle represents a field. The area of circle represents the number of publications. The location of the circle gives an indication of the importance of research done in that field (impact on the y-axis, activity on the x-axis)).

Figure 20: Bibliometric results on all sectors of Swiss research[112]

Switzerland, All Sectors

activity and impact by field 1997- 2001 [total publications *: 17'450]

Fields and Subfields are journal-categories, not categories of institutes or departments



	Fields	Impact	Activity	publ *
f01	Multidisciplinary	132	102	425
f02	Agricultural Sciences	111	68	194
f03	Engineering	115	99	954
f04	Materials Science	127	53	196
f05	Computer Science	104	52	64
f06	Mathematics	102	55	135
f07	Physics	127	124	2586
f08	Astrophysics	119	58	192
f09	Geosciences	109	119	485
f10	Chemistry	121	104	1698
f11	Plant & Animal Science	112	94	733
f12	Biology & Biochemistry	119	101	1191
f13	Ecology & Environment	127	107	293
f14	Microbiology	121	116	383
f15	Molecular Biology & Genetics	120	114	671
f16	Neuroscience	98	108	792
f17	Immunology	135	149	596
f18	Pharmacology	117	101	345
f19	Clinical Medicine	97	109	4834
f20	Psychology & Psychiatry	75	38	215
f21	Social Sciences	78	20	188
f22	Education	23	8	11
f23	Economics & Business	84	43	128
f24	Law	7	37	20
f25	Arts & Humanities	45	17	135

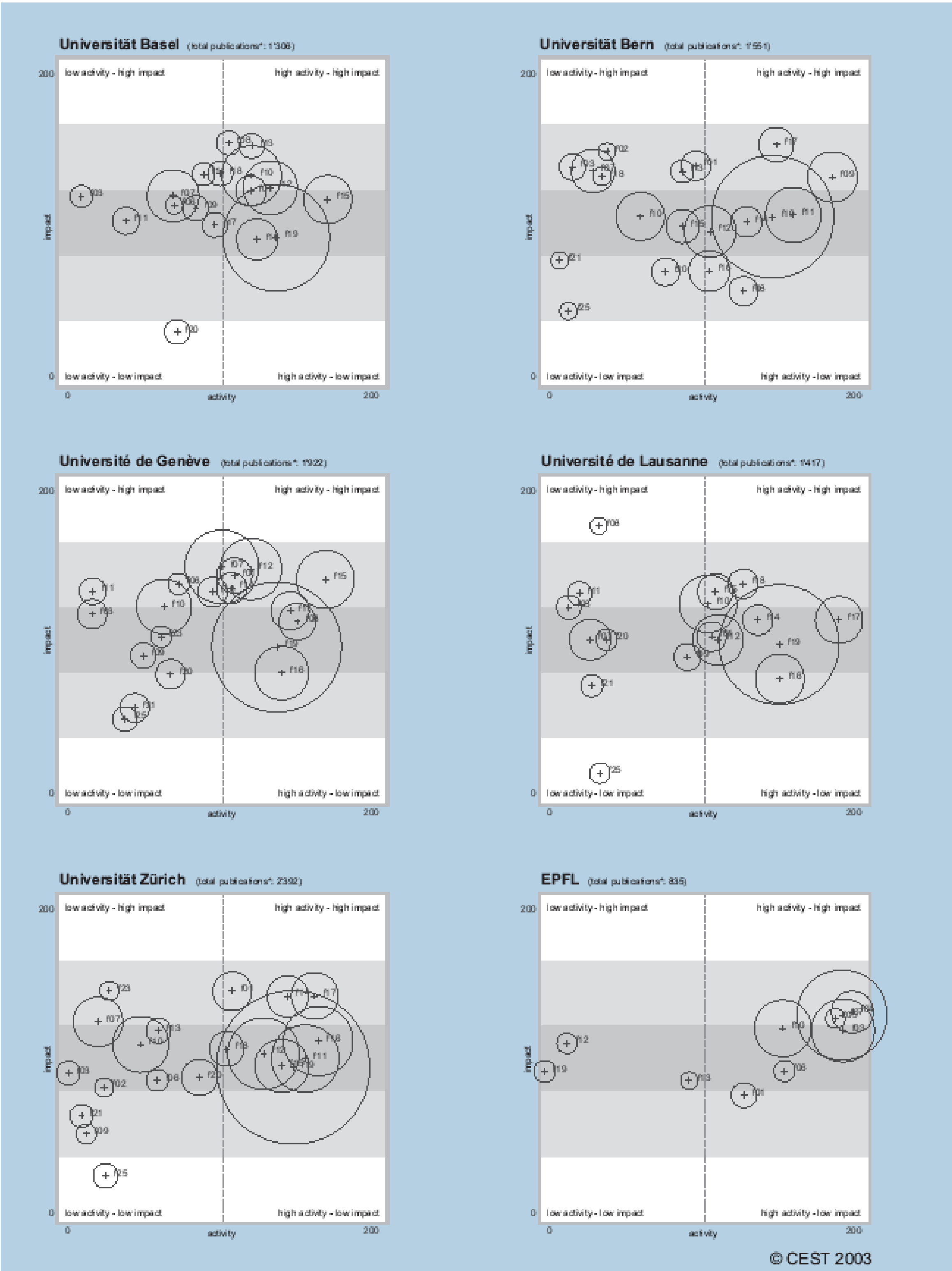
Fields table:
White background:
No proper subfields

Data Source: ISI: SCI/SSCI/AHCI
* Publication numbers: annual mean 1997-2001

[112] pp. 43 of CEST, 2003, *op cit.*

Figure 21 gives the same results as Figure 20 but broken down by each university in Switzerland. Similar results are also available for government research agencies.

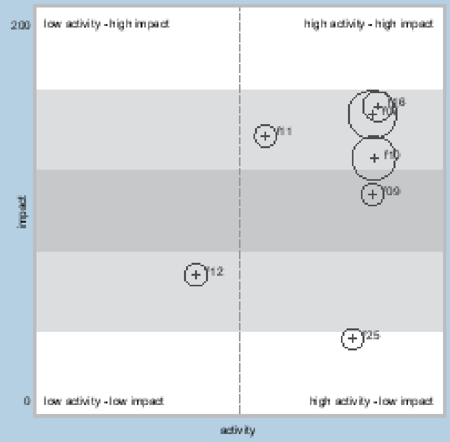
Figure 21: Research output by university in Switzerland[113]



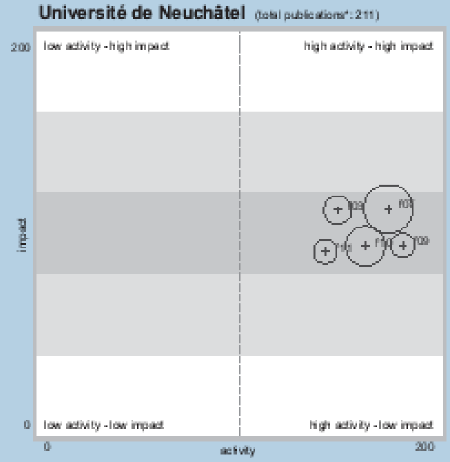
---

[113] pp. 82-83 of CEST, 2003, *op cit.*

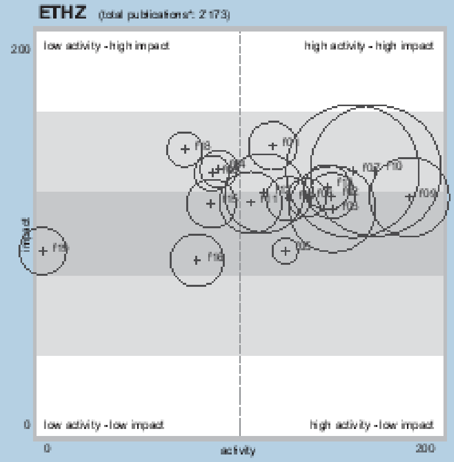
Université de Fribourg (total publications*: 240)



Université de Neuchâtel (total publications*: 211)



ETHZ (total publications*: 2 173)



f01	Multidisciplinary
f02	Agricultural Sciences
f03	Engineering
f04	Materials Science
f05	Computer Science
f06	Mathematics
f07	Physics
f08	Astrophysics
f09	Geosciences
f10	Chemistry
f11	Plant & Animal Science
f12	Biology & Biochemistry
f13	Ecology & Environment
f14	Microbiology
f15	Molecular Biology & Genetics
f16	Neuroscience
f17	Immunology
f18	Pharmacology
f19	Clinical Medicine
f20	Psychology & Psychiatry
f21	Social Sciences
f22	Education
f23	Economics & Business
f24	Law
f25	Arts & Humanities

Fields are journal-categories, not categories of institutes or departments

Data Source: ISI: SCI/SSCI/AHCI

* Publication numbers: annual mean 1997-2001