6 Public funding of academic research: the Research Assessment Exercise of the UK

Jos Koelman and Richard Venniker

6.1 Background

The previous three chapters looked at issues concerning education. In this chapter we turn to research. More specifically, we turn to the funding of research in institutions of higher education. During the last decade competition for research funds and the use of research evaluations have become key issues in technology and science policy in many OECD countries. A major factor behind this trend is the growing demand for accountability of public expenditures, including public research funding, by citizens. Governments and universities are pressed to make more efficient use of public resources, and to give better account of the use of these resources.

In this chapter, we discuss the pros and cons of output-based funding of the research activities of universities. We focus on how it affects the incentives of academic faculty with respect to research, teaching and knowledge transfer. We draw lessons from the UK, which has one of the most output-oriented university research funding systems. Since 1986, research by British universities is evaluated every four or five years in the so-called Research Assessment Exercise (RAE). The results of this exercise play an important role in research funding by the government: low-quality research is not funded at all, and research of high quality is rewarded with relatively generous funding.

In Section 6.2 we discuss the central issues and concepts. A description of the funding and evaluation system of academic research in the UK is given in Section 6.3. In Section 6.4 we discuss the effects of the RAE.

6.2 Research funding and economic theory

6.2.1 Pros and cons of output-based funding

The goal of introducing output-based funding (like the introduction of the RAE in the UK in 1986) is to increase the quantity and / or quality of research output.[1] Whether this increase will come about depends on various factors. Furthermore, introducing output-based funding may also influence activities other than research. In this section we describe the various possible effects, which are listed in Table 6.1.

[1] How to define and measure research output will be discussed in the next section.

Why would one expect a rise in research output? First, reallocation of resources to the most able and productive research groups may raise overall research output. This assumes that the measure of performance that is used accurately reflects marginal research productivity. Little is known about the production function of research, however. Although one of the few robust findings is that the distribution of average research productivity over researchers is very skewed, it is not clear what part of it may be attributed to the ability of researchers (see Stephan, 1996). Second, allocation of research funds between research units (universities, departments, research groups or even individuals) on the basis of performance would strengthen their incentives to provide research effort, and thereby raise their research productivity and eventually aggregate research productivity.

Whether these positive effects of performance-based funding will actually occur depends on several factors. First of all, introduction of explicit incentives for research effort may crowd out intrinsic motivation. Several examples outside the field of science where this crowding-out is supported by the data are described by Frey and Jegen (2000). The relevance for science is unknown.

A second assumption is that individual effort has a positive effect on aggregate research productivity. This relation need not apply due to the tournament character of science. The norm of "priority of discovery" is generally thought to play an important role in academic research: being the discoverer of new (path-breaking) knowledge enhances one's reputation and future research career (see Dasgupta and David, 1994). This importance of being first may give rise to acts of secrecy in the communication of intermediate research results with other researchers (the opponents in the tournament). This possibly tempers a positive effect of explicit incentives on aggregate research productivity.

Table 6.1 Theoretical pros and cons of output-based research funding at universities

Pro	Con
- allocation to (currently) most able researchers	- adverse incentives for non-measurable research effort
- incentives for measurable research effort	- no funds to new, talented researchers
	- crowding-out of intrinsic motivation
	- bias toward low-risk, short-term research and well-established research approaches
	- low comparability of output between scientific disciplines
	- adverse incentives for other faculty activities
	- academic "transfer market"

Whether output-based funding succeeds in raising research quality and / or quantity also depends on the quality of the output measures. Several imperfections of research output measures have been identified in the literature.

When part of the research output is not measurable, funding based on objective indicators may increase measured research output without increasing actual research output. It induces researchers to concentrate their efforts on the measurable outputs of research, which may be detrimental to actual output. Consequently, when non-measurable output is important, weak incentives on measurable output are desirable (Holmstrom and Milgrom, 1991).

Output-based indicators are necessarily based on past research accomplishments which may be misleading with respect to future productivity. Accomplishment-based funding tends to shift the distribution of funds toward older researchers and research units, at the cost of young researchers, re-entering women and new research units that may be more productive in the future, but have had less possibility to express their potential (Lazear, 1997). A similar reasoning applies to new research areas and new approaches versus established ones.

The length of the evaluation period is important as well. Research output is not only the result of effort and ability, but also of chance. Indicators that are based on short evaluation periods may result in one-time luck having long-time consequences due to the so-called Matthew effect[2]: successful research, whether due to ability or good luck, enhances reputation and the chance of obtaining future research funding, and thereby the probability of being successful in the future. Longer evaluation periods mitigate this influence of luck somewhat. A short evaluation period may also distract universities from path-breaking, high-risk research – with results only expected in the long-run – toward short-term and mainstream research with foreseeable output. This runs counter to the accepted view that university research should focus on research that would be under-provided by private parties due to external effects and high uncertainty.

Research is not the only activity of universities. They are also engaged in education and the transfer of research findings to the general public. The incentives on the three activities should be balanced in order to prevent that one of them will be crowded out. When education funding does not depend on education output and the effort academics put into education is hard to verify, strong financial incentives for research may go at the cost of the quality of education. The same applies for the transfer of knowledge, which is a legal task but is hardly rewarded explicitly.

Finally, individual institutions may use intrinsically unproductive strategies to increase their research output. These strategies do not increase the output of the total research system. A possible example that has featured prominently in public debates concerns poaching of researchers from other institutions shortly before an evaluation exercise (especially the timing is unproductive here, since mobility of researchers itself may be very productive).

[2] "To those who have more shall be given", from the Gospel of St. Matthew.

6.2.2 Research output and pitfalls in popular output measures

Evidently, output-based research funding requires a notion of what research output is. In general terms, the output of research is new knowledge. This initially takes the form of tacit knowledge, *i.e.* knowledge in the heads of researchers. Transfer of tacit knowledge requires face-to-face contact, which makes it a relatively expensive affair. To facilitate knowledge transfer, tacit knowledge may be written down on paper or in bits and bytes: it may be codified. Scientific papers, journal articles, patents and computer software are all examples of codified research output.

These codified outputs are the basis of attempts to evaluate the research efforts of universities, research groups and individual researchers. Evaluation of research has been a central component of research activity ever since science is conducted in specialised institutions, beginning in the late eighteenth and early nineteenth century. It has mainly served two types of decisions: funding research proposals and research organisations, and formulating a research strategy.

Various indicators of research output have been developed, all having their pros and cons The two main quantitative indicators are publication counts and citation analysis. Subjective peer review plays an important role as well. The remainder of this section discusses the pros and cons of the different indicators, and is largely based on the overview of international practices toward research assessment by Geuna *et al.* (1999).

The method of publication counts takes the sum of publications produced over a given period as a proxy for research productivity. To account for the quality of publications, different publications may be given different weights. Weights may differ between different types of publications (like books, journal articles, and working papers). Different journal articles may also receive different weights, depending on the journal in which they have been published. One possibility that has been used is to weigh articles according to the journal impact factor, which is the mean citation rate of all the articles contained in the journal, and is published annually in the Science Citation Index Journal Citation Report. Apart from the way quality is taken into account, several other decisions have to be made. Examples are the maximum number of publications that is taken into account, the length of the evaluation period, and the way co-publications are weighed (as a single-authored article, or inversely proportional to the number of authors, or otherwise).

Despite the different refinements of rough counts that have been applied, this performance indicator has several shortcomings as a measure of overall research output:

- Research output other than publications (like patents) is left out;
- The acceptance process for publications may be biased (*e.g.* toward established authors, or toward research within a familiar field or paradigm), and weighting schemes for journals may not be representative for the individual articles (Seglen, 1997);

- The choices about types and number of publications to be included, the weights to be used, the evaluation period, and the way co-publications are treated, are partially arbitrary.

The use of publication counts (and other indicators) for research groups, departments and whole institutions raises three additional issues. First, the proxy should be adjusted for the size of the research unit by taking the number of publications per researcher. Second, the output per researcher for a department may vary considerably depending on the number of staff in a department that is included in the indicator (only senior researchers, also Ph.D. students, maybe all types of faculty). And third, the output of a department may be altered significantly by the mobility of staff. The different manners of ascribing the output of a researcher to a department (based on the affiliation at the time of research, or based on the current affiliation) may have a strong impact on the output indicator.

Citation analysis concerns the counting of the number of times research publications of a researcher are referred to elsewhere in the literature. It is used to assess the quality of the research output. Citation indicators are mostly based on the *Science Citation Index* (SCI) of the Institute for Scientific Information. Besides the shortcomings mentioned above, particularly important shortcomings for citation counts are:

- The SCI tends to have a bias in favour of publications in the English language (and especially towards North American sources), and only the first author is reported in the SCI;
- Citation counts cannot distinguish between positive and negative citations, and may be distorted by citations to academic friends or by self-citations (although the latter are easier to recognise);
- The choice of citation windows (how many years are considered after the publication) is partially arbitrary, and may work out negatively for seminal or radical publications that take some time to be understood, accepted and referred to.

Peer review is the evaluation of research output by peers. Frequently, peers also use quantitative information about publications and citations in their assessments (sometimes referred to as *informed peer review*). In the Netherlands and in the UK, research assessment is mainly based on informed peer review. The most important shortcoming of peer review as a method of research assessment is that it is subjective, and may be insufficiently systematic and transparent. In principle, this may result in:

- Dishonest reporting when peers have a stake in the evaluation outcome;
- A bias in favour of large departments because they are usually better known and contribute to research in a large number of sub-disciplines;
- A bias in favour of a department or researcher at an institute because of the good reputation of the whole institute.

6.2.3 Research funding and the relation with research assessments: international differences

Most countries use a dual support system to fund academic research: both funding of institutions (core funding) and funding of research projects. Countries differ in the extent to which research evaluations play a role. The following approaches to core funding of academic research can be distinguished (based on Geuna *et al.*, 1999):

- (Partial) allocation on the basis of research performance indicators, either directly (Australia, Poland) or via an informed peer review process (*UK,* Hong Kong);
- Allocation on the basis of university size (numbers of students and staff), either completely (Germany, Italy, Norway and Sweden) or in combination with a small part that is based on performance (Denmark and Finland);
- Allocation on the basis of negotiation with the relevant ministry, either without any research evaluation (Austria) or with the use of information from research (and teaching) assessment (France);
- Allocation on the basis of small adjustments to historical patterns (the Netherlands). Although research assessment is carried out, it is not linked to funding decisions.

6.3 The Research Assessment Exercises in the UK

The UK has one of the most advanced research evaluation systems in Europe. Since the middle of the 1980s the UK has had four nation-wide university research evaluations, the so-called Research Assessment Exercises (RAEs), carried out in 1986, 1989, 1992 and 1996. The next RAE is planned for 2001.

The results of the RAE have been used to allocate the research funds by the three UK higher education funding councils (for England, Scotland, and Wales) and by the Department of Education for Northern Ireland.[3] Table 6.2 shows that the funds of these funding councils form a large part of total research funding. The other major funding source concerns the research councils, who allocate funds on the basis of research proposals. The share of the funding councils in total research funding has declined sharply, but they are still the largest single source.

[3] Formally, the assessment of quality (the RAE) and the selective allocation of funds are two separate exercises. But, as McNay (1999) observes, most people outside the funding bodies treated the RAE as covering both the assessment and the allocation of funds. We will use the term RAE mostly in this last sense.

Table 6.2 Sources of research funding for UK higher education institutions (percentage of total funding, unless stated otherwise)

	1984	1991	1997
Funding Councils	58.8	47.8	35.1
Research Councils	17.2	20.3	24.1
Other government departments	7.5	6.4	10.4
UK industry	5.6	6	7
Overseas	n.a.	5.5	8.5
Charities	6.7	11	13.6
Other	n.a.	3	1.3
Total (million pounds)	859	1,989	2,942

Note: n.a. = not available.
Source: HEFCE (2000c).

In this section we describe the method of research funding used by the funding councils and the role of the RAE, the main changes through time, and the results of the 1996 RAE (the last evaluation exercise). Because the funding mechanisms and assessment methods of the four regions of the UK are practically the same, we concentrate on the funding of research carried out by the Higher Education Funding Council for England (HEFCE).

6.3.1 RAE-based funding and overall funding within the HEFCE

The HEFCE provides funds for both research and teaching. Table 6.3 shows the breakdown of the HEFCE-funds in teaching, research and special funding for 1999-2000.

Table 6.3 Breakdown of HEFCE funding in 1999-2000

	million pounds	% of total
Teaching	2,930	69.3
Research	855	20.2
- quality-related research funding	835	19.8
- mainstream	743.3	
- supervision of research students	65.6	
- London extra costs	26.1	
- generic research funding	20	0.5
Special funding	435	10.3
Transitional funding and flexibility margin	10	0.2
Total funding	4,230	100

Source: HEFCE (2000a).

The part of HEFCE-funding that is allocated on the basis of the RAE concerns the quality-related funds, which is almost 98% of the HEFCE research funding. The institutions are free in the internal allocation of the research funds they receive. The allocation of the mainstream quality-related funds between institutions takes place in two stages:[4]

- Allocation of total research funds over the subject areas identified in the RAE;
- Allocation of the funds per subject area over the various institutions.

Both allocations are affected by the quality-rankings resulting from the RAE. We first describe how the quality-rankings are determined, and subsequently turn to the translation of these rankings in funding decisions.

6.3.2 The Research Assessment Exercise of 1996

The quality of research is assessed by (informed) peer review in a Research Assessment Exercise (RAE).[5] In this section we discuss the RAE of 1996, which will inform funding decisions until 2001-02. This RAE involved the assessment of over 55,000 academics from nearly 3,000 departments in 191 institutions (Geuna *et al.*, 1999). Note that since the introduction of a unitary university system in 1992, the UK has no formal distinction between the former polytechnics and related institutions (comparable to the Dutch HBO) and the "traditional" universities (comparable to the Dutch universities). Hence, all institutions of higher education are assessed and funded according to the same rules.

At the beginning of the 1996-exercise, 69 subject areas were defined (called Units of Assessment, UOAs). In each subject area the research output has been assessed by one of the 60 assessment panels of on average six to ten experts. Panel members, some 560 in total, were selected on the basis of nominations by about 1,000 outside bodies (subject associations, learned societies, professional bodies and organisations representing users of research).

Institutions were invited to put forward one application in each subject area.[6] The crucial information for the research assessment has been the research output of the so-called *research active staff*. Institutions were free in the selection of researchers as research active staff. It is important to note that the academic staff that is not submitted as research active does not add to the research volume of institutions as well. Hence, institutions basically face a trade-off between quantity and quality. The 1996-RAE did not assess all the output of the research active staff, but considered up to four works (publications or other forms of assessable and publicly available output).

[4] The two other components of quality-related funding are also determined by the outcomes of the RAE, but in a slightly different way. This is not discussed any further.

[5] Teaching activities are assessed by a separate assessment exercise: the Teaching Quality Assessment (TQA).

[6] Sometimes multiple applications from one institution in one subject area were allowed. Since interdisciplinary research-units may be hard to relate to a single subject area, the RAE sometimes allowed for application in a second subject area. In these cases, a second assessment panel considered the submission as well.

The research assessment resulted in a rating for each research unit (see Table 6.4). These ratings reflect the extent in which research in a unit has achieved levels of national or international excellence. Rating 1 implies "research quality that equates to attainable levels of national excellence in none, or virtually none, of the sub-areas of activity" and rating 5* means "research quality that equates to attainable levels of international excellence in a majority of sub-areas of activity and attainable levels of national excellence in all others" (Geuna *et al.*, 1999). The ratings are thus meant to reflect the level of research quality, and not the position of a department in a research quality tournament where a higher rating of one department automatically means a lower rating for another department. In theory it is possible that all departments receive the highest rating of 5* or the lowest rating of 1.

Table 6.4 Distribution of 1996 RAE-ratings over departments

Rating	1	2	3b	3a	4	5	5*
Research department (% of total)	8.2	16	18.2	14.6	23.2	13.9	5.9

Source: RAE96-database (see www.rae.ac.uk).

The average ratings differ substantially between subject areas. The three lowest average ratings (after translating the rankings to a scale from 1 to 7) are 2.4, 2.8 and 2.8, whereas the scores of the three highest rated subject areas are 5.1, 5.4 and 5.6. The difference between subject areas may reflect true quality differences, but may also be the result of different perceptions by assessment panels of the quality-ratings. These differences in average scores have increased in importance, since from the 1996 RAE onward the allocation of the total budget over the subject areas depends on the quality-ratings (prior to this date the budgets per subject were determined before the assessment).

6.3.3 From RAE-ratings toward allocation of funds

As mentioned before, the allocation of the quality-related research funds proceeds in two stages: allocation of the total funds between the subject areas (Stage 1), and allocation of the funds per subject area between institutions (Stage 2). The RAE-ratings influence the outcome of both stages.

In Stage 1 the total funds are allocated between the different subject areas. The share of each subject area in total funding is proportional to the volume of research in the subject area times the relevant cost weight.

There are three cost weights, reflecting differences in costs of research: for high cost laboratory and clinical subjects (weight 1.7), for intermediate cost subjects (weight 1.3) and for other subjects (weight 1.0).

The volume of research is the weighted sum of five separate components:

- The number of FTE research active academic staff funded from general funds, in departments rated 3b or above, and selected for assessment in the RAE (weight 1);
- The number of FTE research assistants (weight 0.1);
- The number of FTE research fellows (weight 0.1);
- 1.75 times the FTE number of postgraduate research (PGR) students in their second and third year of full-time study, or third to sixth year of part-time study (weight 0.15);[7]
- The average of last two years' research income from charities, divided by 25,000 (weight 0.25). Income from charities is divided by 25,000 (in pounds the average salary of a researcher) to obtain a personal equivalent.

The number of research active staff is the most important measure of volume: it accounts for about two-thirds of the total volume. The volume of research active staff is fixed between two RAEs. The other components of research volume are updated annually.

In Stage 2 the funds per subject area are allocated over the various institutions. For each subject area, the share of an institution in the total funds is proportional to the volume of the research unit it has put forward for assessment in the subject area, times the funding weight of the research unit. The volume of research for each institution in each subject is measured as in Stage 1.

The funding weight follows from the quality-ranking of the research unit determined in the RAE. Table 6.5 shows how the ratings relate to the funding weight. Ratings 1 and 2, which amounts to 24.2 percent of the departments (see Table 6.4), generate no quality related funding. Each rating point between 3b and 5 attracts a weight 50 percent greater than the previous point, while the step from 5 to 5* implies a 20 percent premium.

Table 6.5 RAE ratings and corresponding funding weights

1996 RAE rating	1	2	3b	3a	4	5	5*
Funding weight	0	0	1	1.5	2.25	3.375	4.05

Source: HEFCE (2000a).

6.3.4 Changes in RAE through time and plans for the RAE of 2001

Through the years the HEFCE has continually evaluated and reviewed the research evaluation process and the funding system. This section describes the major changes since the RAE of 1989 (see Table 6.7). The first RAE (of 1986) will not be discussed, since it has been strongly

[7] The multiplier of 1.75 is used to scale the 2 years counted for funding purposes back to a total of 3.5 years, which represents an average period of study for a full-time research degree.

criticised for its lack of transparency and the subsequent changes have been very substantial. In discussing the changes we will follow Table 6.7.

A major change that does not concern features of the RAE itself, but has resulted in debates about the RAE, has been the introduction of the unitary system of higher education in 1992. The formal distinction between the polytechnics and other institutions (comparable to the Dutch HBO) and the "traditional" universities (comparable to the Dutch universities) was abolished, and all the institutions of higher education have subsequently been assessed and funded according to the same rules.

The inclusion of the "new" universities in the RAE has also led to a number of changes in the determination of research output relevant for the RAE:

- Grants for teaching and research were separated. Student numbers were removed from the research funding formula, and research students, research assistants and fellows were included. This change has been structural;
- The choice of which academic staff to include in the research assessment was decentralised toward the institutions. Before 1992 all academic staff was subject to evaluation. Ever since the institutions have been free to put forward so-called research active staff. In this choice institutions face a trade-off between quality and quantity: academic staff that is not submitted as research active does not add to the volume of research as well;
- The relation between ranking and funding was changed (see Table 6.6).

Table 6.6 RAE ratings and corresponding funding weights for 1989 and 1992

RAE rating	1	2	3	4	5
1989 funding weight	0.5	1.5	2.5	3.5	4.5
1992 funding weight	0	1	2	3	4

This change had the important consequence that the lowest ranked units no longer received quality-related funds, whereas previously all units received some funding;

- Basic research and applied research could be evaluated separately. This change has only lasted one period; in the 1996-RAE they were integrated again. The change was inspired by the possibility of an excessive focus of the review panels on output measures that were favourable to basic research (like publications in scientific journals), and thus for the old universities. Separate evaluation proved to be cumbersome and added little to creating a level playing field, and was thus cancelled in the next exercise.

The number of quality categories and the correspondence between quality-ratings and funding weights has been changed several times. The first change, described above, basically introduced a category of institutions receiving no quality-related research funding. In 1996, the number of

quality categories has been increased by two. Basically, both the old category 3 and the old category 5 have been split in two. Furthermore, the lowest two quality categories received no funding from 1996 onward. These changes occurred in response to the general rise in quality rating. Due to this rise the departments that had been able to maintain their position in the top category had nevertheless seen their funding per unit of research volume decline, which was considered undesirable.

Another feature that has been changed several times is the quantity of output that is evaluated. In 1989 there were no rules. In 1992 researchers had to mark with an asterisk the two pieces of output they considered to be best. In 1996, the number of outputs counting for the quality assessment was drastically reduced to four pieces. This change has been made in reaction to the publication explosion following the 1992 RAE. The new arrangement has reduced the incentive to maximise the number of articles by repetition, by lowering the quality standards, or through the breakdown of research into lowest publishable units (Cave *et al.*, 1997).

Until 1996, the distribution of funds between the subject areas did not depend on the quality ratings. Since then, the quality of research no longer only determines the distribution of funds within a subject area, but also influences the distribution of the total budget between the subject areas. As explained earlier in the chapter, the amount of funds allocated to a subject area depends strongly on the research volume, which only takes into account the number of research active staff in departments that exceed a minimum research quality (have a rating of 3b or above). This raises the question of the comparability of quality between different disciplines; a question that is especially interesting given the great spread of average ratings between subject areas. This structure may give assessment panels an incentive to overrate the average quality of research output in order to maximise the share of the own subject area in the total research budget.

Table 6.7 Differences between the subsequent RAEs

	1989	1992	1996	2001
Funding period	90/91-92/93	93/94-96/97	97/98-00/01	01/02 - ...
No. of subject areas	152	72	69	68
University system	binary (55 institutions)	unitary (170 institutions)	unitary (191 institutions)	unitary
Funding of teaching and research separated?	no	yes	yes	yes
Staff assessed	all staff	research active staff (selected by the institutions)	research active staff (selected by the institutions)	research active staff (selected by the institutions)
Funding weight as a function of the quality rating	see Table 6.6	see Table 6.6	see Table 6.5	see Table 6.5
Separate ratings for basic and applied research?	no	yes	no	no
No. of quality categories	5 (see Table 6.6)	5 (see Table 6.6)	7 (see Table 6.5)	7 (see Table 6.5)
Budget per subject area	set before exercise	set before exercise	endogenous	endogenous
Research output per researcher assessed	not specified	two publications + two other output + other research info	best four	best four

Sources: McNay (1999), Williams (1993), www.rae.ac.uk.

The last change, which will be made in the 2001-RAE (and does not appear in the table), concerns the rules for submitting staff that has left an institution shortly before the research evaluation exercise. A "research active researcher" who transfers between two institutions that are eligible to participate in the RAE within the twelve months preceding the census date will be taken into account in the judgement of quality for both institutions, but will only be counted in the research volume of the employing institution at the census date. This change has been made in reaction to references to an academic "transfer market", and should ensure that institutions will not be disadvantaged by staff leaving immediately before the RAE.

6.4 Evaluation of the RAE

What have been the consequences of the RAE? How effective has the RAE been in achieving its goals? And what about unexpected side-effects? This section discusses the impact of the subsequent RAEs. It is mostly based on the main evaluation studies of subsequent exercises:

Williams (1991)[8], Martin and Skea (1992)[9], McNay (1997 and 1999)[10], and the HEFCE Review of Research (HEFCE, 2000b) including the underlying reports.

First a short word about the costs. According to calculations by the HEFCE, an upper limit on the total costs of the 1996-exercise is £37,5 million, just 0,8% of the total funds allocated on the basis of the RAE-results (HEFCE, 2000b).

The total amount of money that changes departments due to the RAE is about 30% of the total funds. Despite these gross money flows, the share of the old pre-1992 universities and the share of the new universities in total funding remains approximately constant (HEFCE, 2000b). The financial consequences for departments may be larger than these figures indicate. This is due to the fact that the RAE-ratings not only determine HEFCE-funding, but also increasingly influence the allocation of other research funds (McNay, 1997). Firms, for example, use the ratings when choosing a research group for contract research or long-time research collaboration.

6.4.1 Research output

The first indications of changes in research output are the changes in quality ratings, reported in Tables 6.8 and 6.9. The changes indicate a steady rise in the quality of research. From 1989 to 1992, 50% of the submissions improved its rating and 35.4% consolidated its rating. The remaining institutions either saw their rating decline or dropped out. From 1992 to 1996 51.7% improved its rating and 31.1% received the same rating in both years.

Table 6.8 RAE grade movements from 1989 to 1992

1992 submissions / 1989 rating	1992 rating 0	1	2	3	4	5	Total
0		90	80	106	47	36	359
1	41	**13**	45	49	10	1	159
2	37	5	**107**	189	58	8	404
3	31	0	46	**284**	176	48	585
4	10	0	1	72	**181**	86	350
5	8	0	1	6	44	**143**	202
Total	127	108	280	706	516	322	**2,059**

Note: Rating 0 indicates "received no rating".
Source: HEFCE (2000b), Annex J.

[8] Williams (1991) interviewed senior staff at sixteen universities.

[9] Martin and Skea (1992) surveyed 117 academics from 25 departments at nine institutions.

[10] McNay (1997) performed a study commissioned by the HEFCE. The study considered the effects of the RAE on the management of research, the quality of research, unintended consequences, the balance between research and teaching, and the nature of research. It involved a literature study, case studies, questionnaires and interviews.

Table 6.9 RAE grade movements from 1992 to 1996

1996 submissions	1996 rating							
1992 rating	0	1	2	3	4	5	5*	Total
0		126	207	131	30	14	7	515
1	180	**78**	84	60	2	0	1	405
2	87	28	**130**	290	44	4	0	583
3	64	2	36	**370**	271	54	5	802
4	13	0	0	79	**254**	162	22	530
5	6	0	0	4	43	**150**	**120**	323
Total	350	234	457	934	644	384	155	**3,158**

Note: Rating 0 indicates "received no rating", rating 3 in 1996 includes 3a and 3b.
Source: HEFCE (2000b), Annex J.

Two questions remain: is the improvement suggested by the increase in ratings real, and can it be attributed to the RAE? The last question is a very difficult one. McNay (1997) emphasises that the effects of the 1992 RAE are hard to disentangle from the effects of other policy changes that took place at the same time: (i) the creation of the unitary system, discussed earlier; (ii) a freeze on the expansion of undergraduate student numbers; (iii) the introduction of teaching quality assessment (TQA), although without significant resource consequences attached; and (iv) more emphasis of government policy on the contribution of academic research to competitiveness and economic strength. Additionally, some rules of the 1992 RAE were changed unexpectedly shortly before the submission deadline.

An international comparison may provide some indications of the efficiency of UK academic research. In 1997, the UK had the largest number of papers and number of citations per dollar (PPP) of higher education R&D expenditures. On the other hand, research funding as a proportion of GDP and the proportion of research funding provided by the government are relatively low internationally (HEFCE, 2000b). At first sight this suggests that UK research does indeed make efficient use of the research resources (although differences in research systems, like the focus in Germany on public research in specialised research institutions instead of in higher education institutions, make firm conclusions difficult).

Additional evidence is provided by surveys of researchers and university administrators. This evidence has the major drawback that it is based on perceptions and opinions, which frequently differ between individuals (even apparently similar ones). Williams (1991), McNay (1997) and Adams *et al.* (2000a) found evidence of improvements in research management: more conscious and transparent planning and monitoring of research, and closure and merger of low-rated departments. Many insiders think research quality has increased, although this is accompanied by more stress among staff.

6.4.2 Funding bias against new researchers

Does the funding system work out negatively for researchers who did not have the chance to prove their abilities in the recent past, like young researchers and re-entering women? HEFCE (2000b) finds no evidence for a bias against young researchers. Two observations support this view. First, the age profile of research active staff submitted to the 1996-RAE is not related to the grade received. Moreover, research-intensive departments even recruit slightly more younger staff relative than the sector overall.

HEFCE (2000b) does find an under-representation of women in the highest-rated departments. The proposed solution is to recognise personal recommendations of peers as evidence in the RAE. When the absence of research output is due to a temporary retreat from the academic labour market, an alternative solution might be to consider the research output in the four years before this retreat.

6.4.3 Bias toward short-term research

McNay (1997) finds indications for several distortions of the nature and content of research. The evaluation period results in more short-term and mainstream research. He also reports a bias of review panels toward more favourable treatment of papers in established scientific journals, leading researchers to focus on more basic research, more mainstream research, and less interdisciplinary research. How severe these distortions have been does not become clear, however. HEFCE (2000b) found no relationship between the percentage of time researchers spend on interdisciplinary research and the rating of their 1996 RAE submission, which suggests the problem is not large.

6.4.4 Adverse incentives for teaching and knowledge transfer

Martin and Skea (1992) report on concerns among academic staff about the negative effect of the RAE on teaching. Jenkins (1995) evaluates the effect of the RAE on teaching in fourteen departments of geography in England and Wales. The paper presents evidence of more teaching by part-timers and postgraduates (particularly in the first postgraduate year), and clear pressures to give priority to research productivity in personnel policy, especially in appointments. Teaching programs tend to become more fragmented, and insufficient new (possibly IT-based) teaching material is developed. McNay (1997) finds similar effects plus a trend toward organisational separation of teaching and research. To what extent these effects influence the educational output remains unclear. Analyses based on proxies for educational output have not been found.

The (negative) effects of the RAE on teaching and the transfer of knowledge are not evident. J M Consulting Ltd (2000) found a widespread view that the RAE did not directly damage the quality of teaching. A negative effect on innovations in teaching, like new teaching material and the attention paid to student support and tutorials, might be present, although views on this are

widely differing. Possibly, a decline in teaching quality may yet have to show up. Even when there is a negative influence of the RAE on teaching and knowledge transfer, the HEFCE-report concludes that this problem should not be tackled by lowering the incentives on research. Rather the answer should be found in attaching greater financial consequences to the Teaching Quality Assessment (TQA), and improving its quality, about which there is much dissatisfaction. A similar argument applies with respect to knowledge transfer.

6.4.5 Academic transfer market

One of the most frequently mentioned aspects of the RAE has probably been the alleged "transfer market" for staff in the run-up to an exercise. The fact that institutions are assessed on the performance of the staff in post on the census date for the RAE has been said (among others in the survey by Williams) to encourage a frenzied transfer market in the period before an exercise. The data do not support this hypothesis. McNay (1997) calculated that only about 1% of total academic staff moved due to the 1992-RAE. The same figure applies to the RAE-related transfers in the two years up to the 1996-RAE, a period in which the entire sector grew by 25%. There has been some timing of retirement in the 1996-RAE: in the year following the RAE, the percentage of staff retiring or moving out of active employment rose from 1.84 percent to 3.30 percent. Mobility may have remained this low because institutions took steps to retain staff, like salary increases, relief from teaching, sabbaticals and provision of support staff.[11] Compared with the US-researchers and with industrial researchers, UK academic (RAE) researchers are relatively immobile (HEFCE, 2000b).

6.4.6 In conclusion

Based on all the above, we arrive at the following summary of the findings concerning the effects of the Research Assessment Exercises in the UK (see Table 6.10).

Table 6.10 Consequences of the UK system of output-based funding, the Research Assessment Exercise

Pro	
Research output (research management)	weakly positive effect
Con	
New researchers / re-entering women	no effect / some negative effect
Short-term, mainstream research	ambiguous
Teaching	weakly negative effect
Knowledge transfer	unknown
Academic transfer market	no negative effect

[11] On the other hand, the percentage of staff moving to another institution in the year after the 1996 RAE was significantly higher than in the two years before the exercise.