Resource Allocation and Research Performance: The Assessment of Research

FINAL REPORT

A study carried out for the Higher Education Funding Council of England

by

Aldo Geuna
Dudi Hidayat
Ben Martin

Brighton, 30th July 1999

SPRU
Science and Technology
Policy Research

Acknowledgements

We would like to thank Satis Arnold, Rémi Barré, Tibor Braun, Linda Butler, David Campbell, Hans-Dieter Daniel, Benoit Godin, Julita Jablecka, Sylvan Katz, Jan Kozlowski, Anne Klemperer, Terttu Lukkonen, Slavo Radosevic, Paul Reynold, Jacky Senker and Štefan Zajac, for helpful suggestions and advice.

CONTENTS

Executive Summary

This report analyses and synthesises the literature dealing with the evaluation of scientific research performance of universities and its relationship with the allocation of research funds from government in the UK and in other European, North American and Asia-Pacific countries. The principal aims of this report were:

1. to produce perhaps the first comprehensive bibliography of academic and policy literature on the assessment of research performance of universities, particularly university research evaluation carried out at the national level, for the purpose of research funding allocation;

2. to provide a critical review of this literature focusing, in particular, on e*valuation performer* and *evaluation purposes* and on *evaluation criteria and methods.*

In most of the 18 countries considered, some sort of 'dual support' system is used; general infrastructural research funds are provided on an institutional basis while specific research costs are met through project grants. Three main approaches to general university funding have been identified.

In the first, university research funds are allocated, at least in part, on the basis of some sort of research evaluation – that is, a performance-based approach. Countries such as the UK, Hong Kong, Australia and Poland are applying performance-based allocations with different degrees of assessment. In the UK and Hong Kong, the approach is based on informed peer review, while in Australia and Poland performance indicators are used.

The second approach is also based on formula allocation, but the measure included in the formula is the size of the teaching and learning side of the university ('education size'). Research funds are allocated together with education funding as part of the general institutional funding to universities. Countries such as Germany, Italy, Sweden and Norway follow this kind of approach. Finland and Denmark allocate the largest share of general research funds on the basis of education size, but a small portion of core research funding is allocated on the basis of the teaching and research performance of the university.

The third model is characterised by a process of negotiation between the universities and the ministry responsible for university funding and through which

general research funds are allocated. In Austria this model is applied without any research evaluation, while in France information gleaned from an evaluation of the university teaching, research and management activities is taken into account during the negotiation.

Finally, there is a fourth category of countries in which a research assessment is carried out but this is not linked with funding decisions. This category can be subdivided into countries where there is a dual-support system (e.g. The Netherlands) and those where there is not (the United States). In The Netherlands evaluations of research performance are used to improve the research quality of the university. Funds are allocated on an incremental and formula-based (student number) approach.

It is important to note that, although only a small number of countries are currently using performance-based approaches to university research funding, most of the countries considered in this report are either in the process of implementing some form of performance-based allocation, or are considering doing so.

This report provides a description of current practices in research assessment at the national level for the purpose of informing the funding allocation process in a large number of countries. It does not provide an analysis of the efficiency and efficacy of the different funding approaches. Moreover, the aim of the study was not to provide a better understanding of research assessment in the UK, nor to make an assessment of possible policy alternatives facing the Higher Education Funding Council for England (HEFCE).

Recommendations/Action Items

- The HEFCE should continue to monitor current practice in university assessment and funding in foreign countries.
- The HEFCE should assess the intended and unintended consequences of the current funding approach used in England, and compare this with one or more of the approaches applied in other countries.

Research Evaluation Practices in the Considered Countries

No	Country	University Research Evaluation for Allocation of Core Research Fund	Other National Evaluation of University Research
1	**UK**	Ex-post informed peer review (RAE)	
2	**Hong Kong**	Ex-post informed peer review (RAE)	
3	**Australia**	Ex-post quantitative evaluation (RQPC)	
4	**Poland**	Peer review + ex-post informed peer evaluation	
5	**The Slovak Republic**	Student numbers corrected by ex-post informed peer review	
6	**Denmark**	Education size*/ Performance based	Research field evaluation
7	**Finland**	Education size/ Ex-post quantitative evaluation	Research field evaluation
8	**New Zealand**	Education size**	
9	**Germany**	Education size	
10	**Italy**	Education size	
11	**Norway**	Education size	Research field evaluation
12	**Sweden**	Education size	Research field evaluation
13	**Hungary**	Education size	
14	**Austria**	Negotiation	Research field evaluation
15	**France**	Negotiation	CNE & CNRS Evaluation
16	**The Netherlands**	Education size	VSNU research evaluation
17	**USA**	Education size	Graduate-Research evaluation
18	**Canada**	Education size	

* That is, the size of the teaching and learning side of the institution.

** It is proposed that from 2000, a portion of the core research funding will be allocated based on ex-ante evaluation: peer review.

1. Introduction

Research evaluation has emerged as a key issue in contemporary science, technology and innovation policy in most of the OECD member countries (OECD, 1997). This has been largely driven by the increasing demand for accountability of research activities in the context of a levelling off or even decline of public resources. The eighties and nineties witnessed a slow growth in the public resources allocated to science, while the number of scientific research activities vying for funding rapidly increased. In this context, governments need to scrutinise every funding allocation so as to 'optimise' their research funding allocation with the aim of 'maximising' the benefits of research for society. In this respect, university research cannot be an exception.[1]

Over the last two decades, society's attitudes and demands upon universities in Europe have undergone great changes (OECD, 1987; Gellert, 1993). In many countries, a growing concern has emerged "… about the increasing cost of funding university-based research … and the need to obtain 'value for money' for public expenditure on higher education" (OECD, 1987, p.19). Universities are expected to become more efficient in using public resources and there is a growing demand for accountability in the use of these resources (Massy, 1996). All these pressures have made university research evaluation a central issue.

It should be pointed out, however, that evaluation is not a new activity in university research management. Since the institutionalisation of science, in the late eighteenth and early nineteenth century, evaluation has been a central component of research activity. Under the broad category of peer review, various forms of research evaluation are implemented, for instance: when research reports are submitted; when papers are submitted for a conference or for publication; when a research council or government funding council allocates a grant for research; or when a new faculty member is appointed. Thus, evaluation takes place on many occasions within university research. Nonetheless, these long-standing evaluation methods are only appropriate for the assessment of a particular project or individual researcher.

Before embarking on further discussion on the evaluation of university research performance, it is useful to define several terms pertaining to evaluation activity. Although a few scholars distinguish evaluation from assessment (see for example Hills and Dale, 1995), for the purpose of this study the terms evaluation and assessment are

[1] An account of the demand from public authorities on universities in Europe can be found in Gellert (1993). For the economic background to the 1986 Research Assessment Exercise in the UK see, among others, Phillimore (1989).

both used to mean looking at an object in terms of a set of quantitative and qualitative performance measures. Evaluation can be distinguished into *ex-ante* and *ex-post* evaluation, and in terms of the function it performs as summative or formative evaluation (Kogan, 1989; Massy, 1996, Suter, 1997). Ex-ante evaluation is evaluation conducted prior to the project execution to assess its feasibility or the likelihood of its success. Ex-post evaluation is conducted after the project is completed to assess its achievement, measured against specific performance measures. Summative evaluation involves judgements about the particular performance of an object based on accumulated evidence in comparison with the performance of similar object. In the case of formative evaluation, however, an evaluator collects and reports data and judgements to assist the development of an object; the results of the evaluation might be used to change its course and better meet its original goals.

The results of research evaluation are increasingly being used as inputs for decision-making. Decisions are made at various levels. It might be the level of programme/project, research organisation or national level – for instance evaluation conducted by a research funding council. Among the various decisions taken based on the result of evaluations, there are two types that are most important: strategy formulation and funds allocation (Steen and Eijffinger, 1998). Evaluation for strategy formulation is generally conducted at the institutional level. The institutions need to evaluate the strengths and weaknesses of their research, evaluate the environment in which they operate in order to prioritise particular research areas and to formulate the strategy for their future. The quality assessment system within universities is an example of such an evaluation practice. However, evaluation for strategy formulation might also be conducted at the national level as part of the effort to strengthen the national research system. In addition, evaluation is used as a tool to decide the allocation of research funding. The research performance of individual researchers, projects, programmes and centres is evaluated and research funds are allocated to the unit with the higher performance. Through research evaluation, it is expected that funds will be allocated to the units that will provide the greatest social or economic return from the investment.

Accordingly, there are two main dimensions to describe evaluation practices: the level of organisation and the purpose of the evaluation, as is depicted in Table 1 (Steen and Eijffinger, 1998).

Table 1. Matrix of evaluation practices

Context/purpose Level	Strategy formulation	Fund Allocation
National/sectional		**X**
Institutional		
Programme/project		

Source: Steen and Eijffinger, 1998

This study focuses on literature relating to the assessment of research performance of universities, particularly university research evaluation carried out at the national level for the purpose of informing decisions on research funding allocation. This sort of evaluation practice takes the form of the Research Assessment Exercise (RAE) in the UK. The Higher Education Funding Council for England (HEFCE) – jointly with the funding councils in Scotland, Wales and the Department for Higher and Further Education, Training and Employment in Northern Ireland – carries out a national evaluation practice every four or five years to determine the quality rating of each research area within universities and colleges. The method of evaluation is mainly based on peer-review. For each research area an assessment panel of on average 6-10 experts is set up. This panel evaluates the quality rating of each research group based on documents submitted by the universities. The documents include among others: publications and other publicly identifiable outputs, the number of research students, external research grants obtained, and the number of academic staff active in research.

This report provides a description of current practices in research assessment at the national level for the purpose of informing the funding allocation process in a large number of countries. Research evaluation approaches developed in the UK and in other countries will be examined. These include a number of countries in Europe – Austria, Denmark, Germany, Finland, France, Hungary, Italy, Netherlands, Norway, Poland, Slovak Republic, Sweden; North America – the USA and Canada; and Asia Pacific countries – Australia, New Zealand and Hong Kong. Research assessment in a format similar to that of the RAE in the UK – with respect to the comprehensiveness and the direct link with the funding allocation mechanism – is only adopted in a limited number of countries. Although the primary concern is research assessment at the national level for the purpose of funding allocation, other forms of research assessment in a number of countries are also reviewed.

2. Methodological Approach

The countries considered use various types of assessment for university research. Here we propose a general scheme to analyse them. The assessment methods of each country have been examined from the following three main aspects. Given the aim of this report (to produce a review of the approaches to assessing and funding research), most emphasis was placed on the first two.

1. *Evaluation performer and evaluation purposes*
   National level:
   a. Research Council
   b. Funding Council
   Institutional level:
   a. Board of University
   b. Department
   Purposes:
   a. Funding allocation/accountability (summative)
   b. Strategy formulation (formative)

2. *Evaluation criteria and methods*
   Criteria: quantity, quality, impact, utility
   Methods:
   a. Peer review
   b. Bibliometrics
   c. Peer review supplemented with bibliometrics

3. *Evaluative remarks*
   a. Strengths and weaknesses
   b. Impact on research and teaching

2.1 Evaluation Performer and Evaluation Purpose

Depending on the country, different agencies carry out the evaluation of research. In the case of the UK, the Higher Education Funding Councils (HEFCs) carry out the research assessment. HEFCs are the funding councils under the Department of Education and Employment with responsibility for allocating teaching and general research funding to universities. Unlike the UK, in the US, independent of the department of education, the Conference Board of Associated Research Councils commissions the National Research Council (NRC) to carry out a comprehensive evaluation of doctoral research

programmes covering all universities in the States (Goldberger *et al.*, 1995). In The Netherlands another type of agency is responsible for the evaluation. In addition to internal assessment by universities, the Association of Netherlands Universities (VSNU) is the main actor in university research evaluation covering all universities in The Netherlands (Steen and Eijffinger, 1998).

Regarding the main purposes of the research evaluation, as mentioned before, two stand out: strategy formulation and funding allocation. The purpose of the evaluation is closely related to who is conducting the evaluation. So, in the UK, HEFCs carry out the evaluation in order to establish the basis for distributing research funding. Within the European context, the United Kingdom is the only country where funding is based on performance (Campbell and Felderer, 1997). VSNU in The Netherlands carries out the evaluation mainly to help the research management of the university, and to support decision making at various levels within universities. However, it should be mentioned that research evaluation also serves the purpose of 'market information'. One of the goals of the Assessment of Doctoral Research Programs in the US, for example, is to present the findings of the assessment in a way that is accessible to educators, administrators, students, and policy makers alike (Goldberger *et al*., 1995). In addition to providing information to decision makers, the assessment also provides students with information on the quality of universities, on the basis of which they can make a better decision about which university is most likely to provide high quality programs.

2.2 Evaluation Criteria and Methods

All research evaluations purport to evaluate research performance. However, different evaluation mechanisms apply different criteria and methodologies. The difference lies in the aspects of the research performance measured. Four main aspects of research performance are usually assessed: quantity, quality, impact, and utility. To measure these aspects of research performance, various indicators have been developed. Before embarking on the discussion on the various criteria and methodologies that might be used in different countries, let us first look briefly at research performance indicators, followed by a discussion of the methodologies that can be adopted.

Research Performance Indicators

A large body of literature has been devoted to the analysis of the difficulties and drawbacks of using indicators of research performance. However, there is no overall agreement on which indicator (or set of indicators) is best suited to measuring the

quantity, impact, quality, and utility of the research.[2] As discussed in Geuna (1999a), the main issues are as follows.

Publication counts allow the evaluation of the research output of individuals, research groups, departments, or institutions. The *per capita* publication count is used as a proxy for research productivity. Among the most commonly cited shortcomings in the use of publication counts are:

1. Depending on the selected type of publication output (journal article, book, review article, etc.), and on the weighting scheme applied, the output indicator may vary considerably (see, for example, Johnes, 1990).
2. The mobility of staff may alter in a significant way the output of a department, consequently different ways of ascribing the output of a researcher to a department – *e.g.* to the one where he or she was based or to the current one – may have an important impact on the output indicator (see, for example, Nederhof and van Raan, 1993).
3. Determination of the number of staff in a department depends on who is classed as a research member of the department. Different accounts for post-doctoral students, Ph.D. students, visiting staff, etc. result in significant variations in the *per capita* figures (see, for example, Hare and Wyatt, 1988).
4. Particularly in medicine and the natural sciences it is common practice to have a large number of co-authors, hence the publications can be counted either on a whole or on a fractional basis, giving rise to different output indicators.
5. The use of publication counts as an indicator of research performance is strongly limited by the fact that the variations in the level of resources (inputs) explains much of the observed difference in publication activity across departments (Johnes, 1992; 1988).
6. Biases favouring the publication of established authors may exist in the publishing process, distorting the significance of the indicator.

The use of publication counts to measure research performance is constrained by the fact that they represent only a measure of quantity and they do not capture the impact, quality and utility of the research.[3] Citation analysis (the count of the citations

[2] For an early work on the difficulties involved in constructing research performance indicators see Martin and Irvine (1983). For references and further discussion see, for example, Cave *et al.* (1997; Chapter 4), and van Raan (1988).

[3] A certain level of quality correction can be introduced into the publication count if the publications are weighted by the impact factor of the journal. For an application of this methodology see, for example, Arora, Gambardella and David (1998).

obtained by a scientist or a department) is used to assess the impact (quality) of the research output. The database most commonly used is the Science Citation Index (SCI) of the Institute for Scientific Information (ISI). Some of the shortcomings referred to above are also relevant to citation counts; particularly important for the latter are the following:[4]

1. The SCI tends to have a bias in favour of publications in the English language and especially towards North American sources.
2. The SCI reports only the first author; moreover it is not uncommon to find programming errors both in the author's name and in the journal citation (see, for example, Moed *et al.*, 1985; Cave *et al.*, 1997).
3. Citations are made not only to works considered to be of high quality, they can also be used in a negative or derogatory way, but citation counts cannot distinguish between the two.
4. Different citation windows (how many years are considered after the publication) may give rise to variations in the indicator measurement.
5. Self-citation, citation to co-authored papers, citations to different journals, all require the development of weighting schemes that at present cannot be applied in an objective way.
6. Seminal or radical works may be difficult to understand or, after their acceptance, become common knowledge, and then may not receive the number of citations that they deserve (see, for example, Cole and Cole, 1972).
7. Citation counts can be distorted by the inappropriate use of the citations such as in the case of a citation circle (researchers unduly citing each others' work) or citations for more personal reasons (e.g. junior staff citing senior researchers).

The difficulties and drawbacks to citation analysis and its costly and time-consuming character indicate that, at the current level of methodological development, and with the available technology, this technique is not optimal for the evaluation of the research performance of large institutions.

The last research performance indicator briefly considered here is peer review.[5] A large body of literature has been devoted to the analysis of the peer review system in

[4] For the analysis of the drawbacks of citation analysis see, among the others, Cozzens (1989).

[5] Research incomes are sometimes considered among the research performance indicators. Nonetheless, as pointed out by among others Hare and Wyatt (1988), research incomes are inputs, and not outputs of the system, and therefore they are not taken into account here.

science.[6] As a performance indicator, peer review is most commonly understood as the evaluation (ranking) of the research output of a department by peers. It allows a quantitative judgement of the research, including analysis of cognitive aspects such as contribution to basic knowledge and contribution to methodology. The three most important shortcomings of peer review as a performance indicator are:

1. Peer judgements tend to be subjective and not sufficiently systematic.
2. Large departments are usually better known and contain active researchers in a large set of sub-disciplines, hence they tend to be favoured in peer review to an extent not supported by quantitative data.
3. The reputation of the whole institution may exercise a positive bias (the halo effect) on the peer review of the department (Fairweather, 1988).

Assessment methods

The foregoing discussion on indicators of research performance suggests that citation analysis and peer review can be used as proxies for research quality. It also suggests that given the shortcomings and drawbacks of citation analysis and its time-consuming and costly character, it is not practical to be employed in evaluation of large-scale activities such as the evaluation of universities at the national level. This renders peer review, despite its shortcomings, as the main method of research quality assessment.[7] Indeed, peer review has played a pivotal role in large-scale research evaluation mechanism in many countries. It often constitutes the core of the research evaluation mechanism, supplemented with information pertaining to publications, citations, and other related information – sometimes referred to as 'informed peer review'. The variation in the method lies in the focus of the assessment and the way the peer review is conducted.

Both in the UK and in The Netherlands research assessment is based mainly on informed peer review. In the UK, for each research area to be assessed, an assessment panel consisting of experts in the area is set up. To incorporate the utility aspect of the research, a user of research, usually from industry, is also invited to sit on the panel. Based on the information provided by the universities and colleges, the panel judges the research quality of a particular research group on a seven-point rating scale. In The Netherlands, a similar method is adopted: for each agreed research area the expert panel is set up and the panel judges the quality of research based on the information provided

[6] For a recent study see Wood (1995).

[7] A survey conducted by Martin and Skea (1992) showed that researchers tend to favour peer review as an assessment method over other methods.

by the university. However, unlike the UK which focuses on a one-dimensional concept of quality, in The Netherlands the quality of research performance is assessed in relation to four different quality dimensions: scientific quality, scientific productivity, scientific relevance and long-term viability. The quality is assessed on a five-point scale ranging from ‘poor’ to ‘excellent’.

2. 3 Evaluative Remarks

No single evaluation mechanism is perfect; there have always been criticisms and at the same time room for improvement. One common feature, until recently, of evaluation practices in many countries has been the low level of transparency of the evaluation process and results. The RAE in the UK is now sufficiently transparent to give opportunities for all relevant parties to exchange opinions in order to improve the exercise. The same case now applies to the evaluation processes in The Netherlands and in other countries. Indeed, transparency is one of the prerequisites for the acceptance of the evaluation results (Helander, 1995; Martin, 1997). It allows for feedback from those involved in the evaluation and makes it possible for the evaluation process to be evaluated and developed in a better way.

When university research is evaluated, the evaluation mechanisms, in one way or another, will exercise control and put pressure on the university research system being evaluated. The evaluation mechanism disturbs the object it purports to evaluate. Therefore, the way in which the evaluation is conducted and its impact on the effectiveness and quality of research deserve scrutiny. In addition, the research evaluation of institutions, such as universities, should be conducted with the full cognisance of the existence of intended and unintended impacts on research and other interrelated functions of higher education, such as teaching and training (Geuna, 1999b; OECD, 1997).

3. Higher education research evaluation and funding in the UK

It has been argued that the UK has one of the most advanced research evaluation systems in Europe (Hills and Dale, 1995), not only evaluation at the level of individual researcher or project, but also evaluation at institutional level, such as the university. Four Research Assessment Exercises (RAE) have been carried out, in 1986, 1989, 1992 and 1996. The next exercise is planned for 2001. The last three exercises had a comparable methodology. By means of wide consultations with the academic community, and between one RAE and the next, the methodology has been continually improved. In fact, one of the strengths of the RAE is its openness to criticism and the willingness of the funding councils as executors to continually modify the procedure in response to the views of the academic community (*Times Higher Education Supplement*, 14 Nov. 1997). In this way, the RAE can be seen as a product of a learning process. The following section discusses the main features of the 1996 RAE in more detail.

3.1 Evaluation Performer and Evaluation Purposes

The 1996 RAE was carried out jointly by the four UK higher education funding bodies – the Higher Education Funding Council for England (HEFCE); the Scottish Higher Education Funding Council (SHEFC); the Higher Education Funding Council for Wales (HEFCW); and the Department of Education for Northern Ireland (DENI) – now superseded by the DHFETE (the Department for Higher and Further Education, Training and Employment).

The RAE aims at giving research in universities and higher education colleges[8] a quality rating on which the distribution of a great deal of public research funds for higher education is based.[9] It completely excludes teaching activities; another assessment exercise, the 'Teaching Quality Assessment (TQA)', assesses teaching quality. The definition of research used for the purpose of the RAE is quite broad. It is defined as (Circular Letter, RAE96 1/94):

> " … original investigation undertaken in order to gain knowledge and understanding. It includes work of direct relevance to the needs of commerce

[8] Since 1993, the higher education system in the UK has had a unified system, in which polytechnics and some major colleges were allowed to apply for the title of 'University' and were given the right to award their own degrees. All universities and higher education colleges are eligible for participation in the RAE.

[9] In the fiscal year 1995-96, in England, no less than 94.3% of the funding from HEFCE for research was allocated based on the results of the RAE (HEFCE, 1997).

and industry, as well as to the public and voluntary sectors; scholarship*; the invention and generation of ideas, images, performances and artefacts including design, where these lead to new or substantially improved insights; and the use of existing knowledge in experimental development to produce new or substantially improved materials, devices, products and processes, including design and construction. It excludes routine testing and analysis of materials, components and processes, e.g. for the maintenance of national standards, as distinct from the development of new analytical techniques."

* Scholarship embraces a spectrum of activities including the development of teaching material; the latter is excluded from the RAE.

There was no separate submission and assessment between basic and applied research. There has been concern that in the science field the RAE has favoured excellence in basic and strategic research to the disadvantage of applied research (Griffith, 1995). In response to this, all panel members were instructed to give equal weight to all research, whether basic or applied research (HEFCE, Circular RAE96 1/94). Panels were only concerned with the quality of work submitted. In accordance with the Government's policy for publicly-funded research set out in the science and technology White Paper 'Realising our Potential', the panel members were instructed to give full recognition to work of direct relevance to the needs of industry, as well as to the public sector.

Furthermore, in response to the feedback from the 1992 RAE suggesting that interdisciplinary research works are not sufficiently rewarded (Royal Society of Chemistry, 1995), universities were encouraged to submit their work of an interdisciplinary nature to the most appropriate assessment panel, allowing for a second related panel to consider the submission. This enabled the evaluating panel to overcome the problem of insufficient expertise in evaluating research of an inter-disciplinary nature by calling for additional expertise from a related panel or from external sources, whenever it was felt to be necessary. In this respect, panels were instructed to give full attention to the special nature of interdisciplinary research and of its assessment. However, the interdisciplinary research continued to be an issue and in the preparation for the next 2001 RAE further attempts have been made by HEFCE to address this problem.

3.2 Evaluation Criteria and Methods

The RAE is basically a research quality assessment by informed peer review, which judges the quality of research based on information supplied by the universities and colleges. Thus, the RAE is an ex-post evaluation. At the beginning of the exercise, all research activities within the institution are structured and subdivided into the smallest unit of research area to be taken as the unit of assessment (UoA) of the evaluation (e.g. biochemistry, biological science, physics). In 1996, a total of 69 UoAs were defined; they were substantially the same as those in the 1992 RAE.

For each UoA an assessment panel of on average six to ten experts was set up. Some 1,000 outside bodies (subject associations, learned societies, professional bodies and organisations representing users of research) were asked to nominate candidates for panel membership. The Chairs of the assessment panels are selected by the Chief Executives of the funding bodies. Half of the Chairs in the 1996 exercise had served as chairs in the previous exercise. The remainders were appointed in the light of recommendations from outgoing Chairs and almost all of them had been panel members in the previous exercise[10]. The Chairs were then asked to make recommendations for the membership of their panels drawn from the nomination of the outside bodies, taking into considerations such things as eminence as individuals, coverage of subject field, and sectoral/geographical balance. In total, some 560 members (including Chairs) were appointed to 60 panels.

Each university or college is free to determine which panel it will submit the required information (one panel for each subject area). Every department is normally assigned to only one UoA. The department forwards the information on its research performance to the panel selected. This information in the case of the 1996 exercise, included (HEFCE, Circular RAE96 1/94):

- An overall staff summary. This contains information on all academic staff and research support staff, whether or not they are included as research active staff.
- Details on the research active staff. Research active staff (in post on 31 March 1996) are those whose work was to be evaluated.
- Publications and other public output. For each active member of the research staff up to four items (either publications or other public output) could be submitted. The

[10] It should be noted that for the 2001 RAE, in response to the concern that 'rolling-forward' two third of panel members from one RAE to the next does not represent dynamism, the panel members will only be allowed to serve for two successive exercises (HEFCE, Circular , 3/98).

work must have been produced during the period January 1990 to 31 March 1996 (six years) in the case of arts and humanities subjects; and the period 1 January 1992 to 31 March 1996 (four years) in the case of all other subjects.

- Overview of research students and research studentship.
- External research income, including the amount and the source of research income.
- Statement of research plans.

Of all this information, the publications and other research outputs form the core of material for the assessment. Unlike the 1992 RAE, which in addition to the submission of two publications, required the total *number* of publications to be submitted, the 1996 RAE required only up to four research outputs per research active staff to be evaluated.[11] This change was introduced to try to assess the quality of the research output rather than its quantity. In addition, this arrangement got rid of the problem of low quality 'rush publication' as a result of the pressure to produce as many publications as possible to increase the total number. There is some evidence that the use of 'total number publications' as a performance measure led to 'publishing inflation' – maximisation of the number of articles produced either by repetition, the lowering of quality standards, or the breakdown of research into lowest publishable units (Cave *et al.*, 1997).

Other important changes were included in the 1996 exercise. The following are the most important:

a. A crucial improvement in the 1996 exercise was the setting and publication of criteria for assessment in advance of the submissions being finalised. This satisfied the transparency requirement. Funding bodies asked a number of panel members from the previous exercise to set up the 'General Guidance for Criteria Setting', which was then published by funding bodies. With reference to this general guidance, the assessment panels were required to produce statements of their individual criteria for assessment, and some account of their working methods, to be published in time to inform the submissions.
b. The period from which published work could be selected was amended from a standard four years to allow the submission of work in humanities and arts subjects produced over a six year period.

[11] The department is not obliged to submit the research outputs of all the staff; it can decide to submit only research active staff, the implication being that it will receive funds only for those researchers.

c. The provision for making separate assessments in the same subject in 'basic/strategic' and in 'applied' research was abolished.

On the basis of the information supplied, the assessment panel judged the research quality of the university department and assigned the rating point. The rating scale adopted is as follows:

Rating point 5*: 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.

Rating point 5: Research quality that equates to attainable levels of international excellence in some sub-areas of activity and to attainable levels of national excellence in virtually all others.

Rating point 4: Research quality that equates to attainable levels of national excellence in virtually all sub-areas of activity, possibly showing some evidence of international excellence, or to international level in some and at least national level in a majority.

Rating point 3a: Research quality that equates to attainable levels of national excellence in a substantial majority of the sub-areas of activity, or to international level in some and to national level in others together comprising a majority.

Rating point 3b: Research quality that equates to attainable levels of national excellence in the majority of sub-areas of activity.

Rating point 2: Research quality that equates to attainable levels of national excellence in up to half the sub-areas of activity.

Rating point 1: Research quality that equates to attainable levels of national excellence in none, or virtually none, of the sub-areas of activity.

The resulting rating point is used in the formula to determine the level of research funding for a unit of assessment, the basic fund allocation per point for that unit being determined in advance. The total block grant then received by an institution is aggregated from calculations for each unit of assessment in which research active staff have been evaluated. The four funding councils use different formulae to arrive at a block grant. In England, each rating point corresponds to a funding rating point, which is then used in the funding formula. For assessment ratings of 5*, 5, 4, 3a, 3b, 2, and 1 the funding rating points are 4.05, 3.375, 2.25, 1.5, 1.0, 0 and 0 respectively. Those departments with assessment rating points of 1 and 2 get funding at the rating point of 0,

which means that these departments do not receive any funding. The funding formula is (see, for more detail, HEFCE, Circular 4/97):

£ = [q + 0.15m + 0.10n + 0.25c] * r * p

£ is the total sum allocated to an institution for a particular unit of assessment

q is the number of research active staff members in that unit

m is number of research students

n is number of research staff (research assistants and fellows)

c is the amount of research income obtained from charitable resources divided by £25,000.

p is the basic allocation per point for that unit

r is the funding rating point (4.05, 3.375, 2.25, 1.5, 1.0, 0 and 0)

This ensures an explicit and formalised system of higher education research funding through the rating points resulting from the research assessment performed by informed peer review.

The publication of the result of the 1996 RAE was followed by a league table compiled by the *Times Higher Education Supplement (THES)* for all institutions in England (*THES*, 20 December 1996). The table is based on the rating points. The rating point is converted to a numerical scale from 1-7. The rating points 1, 2, 3a, 3b, 4, 5, 5* are converted to 1, 2, 3, 4, 5, 6, 7. The composite rating is arrived at by multiplying the rating point with the number of research active staff in a given UoA and taking the average grade for all research active staff in the university through following formula (Tomlin, 1998; *THES*, 20 December 1996, p. ii):

$$\text{CompR} = \frac{\sum_{\text{UoA1}}^{\text{UoA69}} (\text{RAS} * \text{Rating})}{\sum_{\text{UoA1}}^{\text{UoA69}} \text{RAS}}$$

where, CompR = Composite Rating

RAS = Research Active Staff

Rating = RAE grade on seven point scale

Σ indicates the sum of all UoA

The maximum possible composite rating was 7. For the 1996 RAE the THES table ranked Oxford University first with a composite rating of 6.67, followed by

Cambridge and The London School of Economics and Political Science with 6.49 and 6.27 respectively.

The league table, however, should be interpreted cautiously. Tomlin (1998) has raised two main concerns about the meaning and use of the league table. One concern pertains to the fact that the rating is based on the evaluation of research active staff only, and not on the whole staff of a department. Other staff, whose works are considered by a department as being below the required standard, will not be put forward for assessment. Thus, a 'proper' strategy to determine which staff are to be submitted as research active staff would give an institution a higher position in the league table compared to one with the same average quality of staff but which included more staff in its submission to the RAE for assessment.

Another concern originates in the fact that no university or college is the same as another. Each has its own specialisation and, therefore, differs in terms of the UoA that it submits. It should be recognised that the average rating for different UoA varies significantly. The average for the Biochemistry UoA was 5.9 and for Metallurgy and Materials it was 5.3; while for Art and Design it was only 3.9 and for Nursing it was as low as 2.6 (*THES*, 20 December 1996). Therefore, University A comprising relatively successful departments of Arts and Nursing which have been graded at 4 (higher than the average of 3.9 and 2.6 for Art and Design UoA and Nursing UoA respectively) would stand side-by-side with University B which has relatively unsuccessful departments of Biochemistry and Materials which have been graded at 4 (lower than the average of 5.9 and 5.3 for Biochemistry and Material respectively). The problem here is that the league table is not comparing like with like. University A and University B are not comparable, as they include different departments with different standards of quality. Nevertheless, in the league tables, they are compared as similar institutions.[12]

3.3 Evaluative Remarks

The 1996 RAE was the biggest university research evaluation practice of all time. It involved the assessment of over 55,000 academics from nearly 3,000 departments in 191 institutions (*THES*, 20 December 1996). Nevertheless, it was relatively inexpensive. The total cost of the whole exercise has been estimated at £27.3 million, which represents less than 1% of the funds to be allocated (HEFCE, Circular 4/97).

[12] To overcome this problem Tomlin (1998) has proposed another formula to determine the composite rating, which turns out to deliver significantly different rankings for some institutions to those in the THES league tables.

Nonetheless, it is crucial to acknowledge that this estimation of the costs does not include the indirect costs incurred by the departments in organising and managing the submission.. Martin (1997) suggested that the cost of the evaluation process should not exceed 1% of the total research funding being assessed.

Evaluating the RAE requires looking back at the history of the exercise and how it has developed. It has been suggested that the first RAE, conducted in 1986 by the University Grants Committee (UGC), which was succeeded by the UFC and the HEFCs, was probably the first attempt in any country to assess comprehensively the quality of higher education research at the national level (UFC, 1989). Not surprisingly, it was criticised by many scholars.[13] However, it should be noted that most critiques are not so much directed towards the evaluation, but to the methods used by the UGC (Smith, 1987). Following this first exercise, there were attempts by the UGC to review the criticisms comprehensively and to take them into account in the planning of the 1989 RAE. This has now become routine: from one RAE to the next, the funding councils conduct a wide consultation with the academic community and others with an interest in the RAE.

A debate on the advantages and drawbacks of the RAE as a tool to assess and allocate funds to higher education research is currently ongoing.[14] An in-depth discussion on this matter is beyond the scope of this study. The impact of the RAE on teaching is discussed briefly below.

The aim of the RAE is to provide a quality assessment of the research conducted in each academic subject in all government-funded higher education institutions; in turn, this quality assessment will be used as the basis for decisions on research funding allocation. As such, it is used as an instrument to allocate selectively the research funding based on excellence: the better the quality of its research, the more research funding an institution receives. Nevertheless, the RAE has a wider impact than has been supposed as a selectivity policy instrument (McNay, 1999). It is clear, for instance, that the RAE has persuaded some institutions to prioritise research and make research productivity one of the main requirements for staff recruitment (Ball, 1997).

[13] For a compilation of criticism see, among others, Phillimore (1989) and Griffith (1995). The statistical analysis of the RAE can be found in Johnes and Taylor (1992) and Johnes *et.al.* (1993)

[14] See, among others, Bourke (1997), Cooper and Otley (1998), El-Khawas and Massy (1996), Kogan (1998), Kushner (1996), Geuna (1999a; 1999b), *THES* (28 February 1997) and Whittington (1997).

Among the studies on the impact of the 1992 RAE, that by McNay (1999) was the most extensive. As to the impact of RAE on teaching, the McNay study uncovered evidence that there were structural changes within institutions that might jeopardise the quality of teaching. The research staff were expected to produce quality research and therefore were freed from some of their teaching responsibilities, leaving their department to organise undergraduate teaching. Of the heads of unit interviewed, 71 per cent said this had a positive impact on research; 62 per cent said there was a negative impact on teaching.

The negative impact on teaching was in accordance with the study conducted by Jenkins (1995). From his survey of a number of university geography departments in England and Wales, Jenkins highlighted two main problems with regard to teaching. First, as a result of the higher reward for research than teaching and the corresponding high concentration of full-time staff on research activities, an increasing proportion of teaching activity was being undertaken by poorly paid and inadequately supported postgraduates and part-time staff. Secondly, in the face of research pressures there was a reluctance among the staff to devote adequate time to developing teaching materials, such as textbooks and computer-supported teaching material. Jenkins produced evidence that, especially in departments with high research ratings, staff were being withdrawn from expensive teaching and learning technology programmes and that lecturers were being diverted from writing textbooks. These trends endanger the quality of teaching, especially first-year teaching that is pivotal to delivering good results.

4. University Research Evaluation and Funding in Europe

The economic and social pressures raising the need for effective and selective fund allocation in the UK are equally present in other countries. However, evaluation practices at the national level for the purpose of fund allocation, such as the RAE, are only adopted by a limited number of countries. In Europe, the Slovak Republic and Poland are the only two countries that directly link the results of university research evaluation with research funding allocation.[15] Although The Netherlands carries out research evaluation at the national level, using an informed peer review method similar to the RAE, it does not base research funding allocation on the results of this evaluation. In the following we describe university research evaluation and funding systems used in 12 European countries.

4.1 The Netherlands

University research in The Netherlands is financed through a dual-support system. One component is government institutional funding, which can be classified as core funding. Universities traditionally have relied on government institutional funding provided by the Ministry of Education and Science through the so-called 'first-flow' of finance. The other side to a dual-support system is complementary research funding consisting of the 'second flow' and the 'third flow' of finance. The former are project and programme grants provided by research councils and foundations. The latter are contracts awarded to universities by government departments, agencies and various other organisations.

As in other countries in the late 1970s, growing concern about the quality and the social relevance of university research has raised the need for accountability. As a result, a government White Paper (MES, 1979) on University Research published in 1979 recommended changes in the management of academic research. As a consequence, a system of 'conditional funding' was introduced in 1983 (Irvine, Martin and Isard, 1990). In this model, a clear distinction was drawn between teaching and research funds, termed, respectively, A-part and B-part. In the research part, positions for academic staff are financed on the basis of proven quality of research. In addition, a national research objective was determined with the intention that an increasing proportion of institutional funding would be earmarked for research in particular agreed fields in accordance with national research objectives. For this purpose, universities are

[15] Since 1997, in Portugal the Fundaçao para a Ciência e a Tecnologia (Science and Technology Foundation) has been responsible for the funding and evaluation of scientific research. University research funding is based on a three-year budget system that is related to the evaluation of university research performance. Detailed information in English on evaluation criteria and methods is not available.

required to produce a medium-term research plan identifying areas of research in which they have relative strengths and, therefore, deserve priority funding. This 'conditional funding' scheme was in place for ten years, until in 1993 it was succeeded by the HOBEK funding model, which was still used up to 1999.

The HOBEK funding model consists of three parts, namely teaching (23%), research (64%) and 'interweavement' (13%). This funding mechanism is based on a formula with four funding bases: number of registered students, degrees, doctorates and designers' certificates. A large part of the research budget has been allocated on the basis of historical data rather than norms or quality (Koelman, 1998). This means that an incremental funding mechanism has been adopted. In early 1999, this model was replaced by a new model called STABEK2, standing for stable funding (STAbiele BEKostiging), which actually was first proposed in the 1996 Higher Education Plan 'HOOP'. The essence of this model is that government approves funding to universities covering several years, this longer period of time providing universities with more stable funding. This scheme, however, is intended only to be temporary while the Ministry of Education, Culture and Science is developing a new funding scheme in which performance is much more emphasised (Meer, 1999). Thus far, there has been no institutional evaluation of research performance for funding allocation.

Evaluation performer and evaluation purposes

It is clear from the foregoing discussion that in The Netherlands there has been no research evaluation carried out for the purpose of funding allocation. However, university research evaluation has been conducted for strategic formulation.

Rip and Meulen (1995) argued that in the context of evaluation culture in The Netherlands, informal assessment and bottom-up evaluation activities are dominant and that science policy agencies have been interested in strategic changes in the research system, rather than in evaluation. In accordance with this observation, the Association of The Netherlands Universities (VSNU) has been very active in initiating a systematic evaluation of university research. In 1992, 13 universities[16] and the Minister of Education reached an agreement that the VSNU should develop a system of external research evaluation as a complement to the internal efforts on quality control. The evaluation procedure was implemented for the first time in 1993-94, and a complete

[16] The Dutch university system consists of 14 universities – nine general universities, three technical universities, one agricultural university and an Open University. However, due to the different nature of the Open University, it is not included in the evaluation.

first round of evaluation covering all the 27 disciplines was completed in 1998. A second round started in 1999.

The primary objective of the evaluation system, which is officially called Quality Assessment of Research, is to help the research management of the universities to develop their research strategies, and to supply information necessary for decision-making at the various levels within universities, up to that of the executive board. This evaluation system replaced the system of conditional funding that was introduced into the university system in 1982-83, whose objective was to make university research more efficient, productive and socially relevant (Steen and Eijffinger, 1998).

Evaluation criteria and methods

The research assessment practice in The Netherlands covers all academic research conducted within universities. For the purpose of the evaluation, the whole spectrum of academic research is classified into several disciplines that, in turn, consist of several research programmes, which represent the smallest unit of assessment. Hierarchically, research programmes are organised under university departments. Each discipline is evaluated by a separate committee. For instance, the Review Committee for Chemical Research evaluates all research programmes within the chemical research field. In total, there are 27 disciplines, less than the half of the 69 units of assessment in the UK RAE.

Unlike the RAE in the UK, which evaluates all disciplines simultaneously within a certain period before a new fiscal year, the 27 disciplines in The Netherlands are not evaluated simultaneously. This is not necessary because research evaluation in The Netherlands is not intended for funding allocation. It is envisaged that every discipline will have been evaluated within a time frame of four to six years.

In 1993, a first small group of disciplines was evaluated. This first evaluation was used as pilot for succeeding ones. Based on this experience an improved official guideline for evaluation was set up in 1994 – the ‘Protocol 1994 for Quality Assessment of Research’. This bilingual document (Dutch and English) contains important procedural details for the evaluation and sets out the responsibilities of all the actors involved. It also standardises the translation of terminology used from Dutch to English. In the following, the evaluation procedure according to the protocol in 1994 is described.

To evaluate each discipline, a Review Committee consisting of five to seven experts with research and management experience is set up by VSNU in consultation

with the deans of the faculties involved in the area of knowledge to be evaluated, agreed by the Royal Academy of Arts and Sciences (KNAW). The committee chair must be either a Dutch expert or a foreign expert acquainted thoroughly with the Dutch situation. It should be pointed out that, as an important characteristic of the committee, the members of the committee (four to six, as a rule) are predominantly foreign experts. This assures a high degree of impartiality of the committee. Given the international nature of the committee, the primary language of communication during the evaluation procedure is English, and the results of the evaluation are also compiled and published in English.

The committee evaluates the performance of the research programmes based on the information provided by related faculties, which covers research activities during the last five years. The information includes:

- an overview of academic staff;
- a summary of the programme mission and research plan;
- content of the programme and its main results;
- a list of all publications;
- list of five selected key publications from the programme;
- other indicators of quality and reputation (such as patents, invited lectures, etc.).

In addition, interviews with the research programme leaders (who as a rule, are full professors) and site visits are often conducted to supplement the written information. Also, whenever available, a comprehensive bibliometric analysis is provided to the Committee as additional information.[17] The Committee considers the bibliometric analysis to be a useful tool that complements the written and oral information on which the assessment procedures are based (VSNU, 1996).

The Committee assesses the research performance of the individual research programme in relation to four aspects:

1. **Scientific quality:** In the assessment of quality, attention is focused on quality measures, such as originality of ideas and methodology, the importance of research

[17] For example, the Review Committee for Chemical Research was provided with a bibliometric analysis produced by the Centre for Science and Technology Studies at Leiden University (CWTS). The analysis pertained to the impact of publications in scientific journals in relation to all research programmes assessed by the committee.

output for the performance of a discipline, the scientific impact of the research activity and the international prominence of the research group.

2. **Scientific productivity:** This aspect relates the inputs to the outputs of research activities. Number of staff and the size of the monetary resources allocated to research are considered to be the measures of input. Important indicators for outputs are the number and nature of publications in scientific journals and books, the number of dissertations, patents and invited lectures.
3. **Scientific relevance:** For this aspect, the research is assessed in terms of its relevance to the advancement of knowledge in the discipline and science in general, and the possible impact and application for future technology as well. In addition, the benefits to society are also considered.
4. **Long-term viability:** This aspect is assessed based on the submitted plans and ideas for future research. In addition, the publication policy of the research group, the coherence of the programme and the continuity of research lines are also assessed.

Although for the 1999-2002 research assessment the framework remains basically the same, in the new round there is a greater possibility for assessing groups in relation to their own stated (and differing) missions. This means that assessment committees can choose to evaluate a group in terms of their relevance and viability (two of the four basic elements that are assessed) according to the group's own mission. Quality and productivity (the other two elements) will continue to be assessed according to one standard, as in the previous round. The basic change is that some room for flexibility and recognition of differences has been built into the system, and the VSNU has accepted that there are differences between institutions. However, it remains to be seen to what extent this flexibility will be exploited by the assessment committees. The committees are free to decide for themselves to what extent they will take into consideration differing group missions (personal communication with Anne Klemperer).

The assessment of the four aspects is translated into the five-point ratings ranked as : 1 'poor'; 2 'unsatisfactory'; 3 'satisfactory'; 4 'good' and 5 'excellent'. The exact interpretation of these ratings might differ slightly from one committee to another. The following ratings with their corresponding interpretation are used by the Review Committee for Chemical Research.

The criteria for quality (Q) are:

Q = 5 Excellent :	The quality of the programme is equal to that of the world's best in the field
Q = 4 Good :	The quality is above the world's average for that field and may contain elements that are excellent in quality
Q = 3 Satisfactory :	The quality of the programme is comparable to the world's average in that field, and may contain aspects that are important nationally
Q = 2 Unsatisfactory :	The quality of the programme is below the world's average in that field.
Q = 1 Poor :	The quality is far below the world's average in that field.

The criteria for productivity (P) are:

P = 5 Excellent :	The productivity is comparable to the productivity of the world's best groups in that field
P = 4 Good :	The productivity is above the world's average
P = 3 Satisfactory :	The productivity is comparable to the world's average
P = 2 Unsatisfactory :	The productivity is below the world's average
P = 1 Poor :	The productivity is far below the world's average

The criteria for relevance (R) are:

R = 5 Excellent :	Notable and influential contributions have been made to prominent fields. The research group plays an important role in relevant scientific communities.
R = 4 Good :	Either less influential contributions have been made to prominent fields or else influential contributions have been made to less prominent fields.
R = 3 Satisfactory :	The group has performed moderately well in not very prominent but useful sub-fields.
R = 2 Unsatisfactory :	The research is not of great relevance to the actual or potential advancement of knowledge in chemistry, and has little impact on future technologies.
R = 1 Poor :	The research is of no relevance

The criteria for long-term viability (V) are:

V = 5 Excellent :	The group is judged to have clear and coherent plans in line with international trends in the field. It has achieved, and is likely to keep, a distinguished position in international networks. Continuity of funding is guaranteed. Highly qualified staff will continue to be available.
V = 4 Good :	The group is seen as competent to do the proposed research. The scientific issues being researched seem to be fruitful. The research group's position in the field is guaranteed. There are no doubts about future funding or the availability of competent staff needed to maintain the programme of research.
V = 3 Satisfactory :	There is some reservation about one or more of the aspects mentioned above, but there is a reasonable chance that the group will survive and will continue to contribute adequately to its scientific field.
V = 2 Unsatisfactory :	Based on the plans presented, the Committee has serious doubts about the continued viability of the research group, and considers that, without additional measures for strengthening the group, it will not be able to function adequately.
V = 1 Poor :	For one or more reasons, the research programme is judged to be non-viable, and should not be continued.

Evaluative remarks

In 1996, the VSNU initiated a study to monitor the follow-up of the assessment (Westerheijden, 1997). The study focused on the use of the evaluation process and evaluation results by managers in faculties and at the central university level. In addition, other effects related to researcher behaviour and administration, resulting from the evaluation process as well as from the evaluation outcomes, were also studied. According to this study, the follow-up by universities of the assessment has been satisfactory. The universities have been using the results of the assessment either actively (using them as a basis for policy choices) or passively (the reports are being read and discussed). Active use seemed to prevail: based on the assessment results, nearly all faculties have drawn up a strategic action plan to implement improvements.

Another important finding of this study was the sense of 'external legitimisation' of an already established reputation, which gives an administrator a justification to manage research in a particular direction based on the evaluation results. According to the respondents, the VSNU research evaluations did not often bring completely new information to light, a view reflected in statements such as: "External judgements legitimise what you, as an administrator, know". The degree of legitimisation depends on the credibility of the evaluation, which in turn depends on a number of factors. Several important factors emerging from the interviews in this respect are (Westerheijden, 1997 p. 405):

1. The reputation of the evaluators and the composition of the evaluation committee,
2. The fact that these evaluations cover practically all the research of the evaluated discipline,
3. The differentiated judgements (on a four- or five-point scale), and
4. The fact that the evaluation results are made public.

With regard to the effects of the evaluation process as well as of the evaluation results, the study identified the following types of effects:

1. *Co-operation among researchers*. It was concluded that the "interdependence of researchers has grown during the last ten to fifteen years, but that the VSNU evaluations do not play an important role in this respect, because other policies affect this area".
2. *Publication behaviour*. Evaluations have put pressure on researchers to publish, and to publish ever more 'strategically' in international (English-language) journals with high impact factors according to the ISI indices. As a result, researchers' publication behaviour has become more and more consciously strategic.
3. *Relationships between and among university actors.* As a result of external legitimisation, the study identified the increasingly important role of administrators in university management.
4. *Evaluations and funding allocation.* A formal direct link between evaluation results and funding allocation does not exist in The Netherlands, and is not desired by respondents. The main reason for this, as is argued by respondents, is that the link might lead to a complete loss of research funds for a faculty and, thereby, would undermine the Humboldtian unity of research and teaching, threatening the distinction between universities and colleges (Hoogescholen).

5. *Position vis-à-vis external actors*. Universities do not actively disseminate evaluation results outside of their institutions. It seems that universities do not actively try to influence individual external actors through the reports. Nevertheless, a 'halo effect' does exists, affecting professors' reputations in the eyes of external actors, which in turn influences the earning capacity for contract research, consultation, etc.
6. *Improvement in quality?* The study cannot answer this question definitively. However, external evaluation, such as the VSNU evaluation, seems to exert a push on researchers towards doing 'better' research, as reflected in the words of one respondent: "Research leaders are thinking about how they can score better next time, because some were disappointed at being not 'excellent' but just 'good'. Others were relieved by receiving this same score".

It is worth mentioning at this point, with regard to the general impacts of the assessment, that in an interview carried out by Felderer and Campbell (1998) the majority of Dutch experts held the following opinions:

- The assessment has increased the quality of research and also awareness among researchers of the importance of research quality has been increased. However, it should be read with caution with reference to the above VSNU study (point 6).
- Driven by the competitive environment, the research management within universities has been improved.
- Only a few of the assessment results were surprising. The majority were as predicted informally by the academic community.

Finally, it should be pointed out that research groups are given an opportunity to appeal formally (although there have been few instances of such appeals) against the result of the evaluation, but there is no independent last resort appeal body (Steen and Eijffinger, 1998). This has probably awarded more credibility to the evaluation mechanism and increased the acceptance of the results.

4.2 France[18]

The higher education system in France comprises 160 institutions. Research activities take place mainly in the universities (approximately 80) and Grandes Écoles, which comprises 28 engineering schools, 27 management schools and 14 specialised schools.

[18] This section is based on Senker (1999), Laredo (1998) and personal communication with Rémi Barré.

Public funding for research in Grandes Écoles is provided for the most part (34% in 1994) by the responsible ministries. In addition, funding from contracts with industry and EU programmes contributes 38.5% to the total research funding. Research in universities, on the other hand, receives financial support from the Ministry of Education, Research and Technology (MNERT) in the form of core funds that cover salaries for full-time technicians and support staff, grants for PhD students, support for infrastructure, such as equipment and buildings, and resources that the university can allocate according to its particular strategy. Salaries for professors and lecturers are provided by other MNERT streams of funds. As professors and lecturers are expected to spend half of their time on research, these salary funds can be perceived as contributing to the financial support for university research. In addition, the regional government also provides funds for university research.

In the French research system, national research activities are not funded through research councils, as is the case in many countries. Research is carried out by various mission-oriented research organisations, and the research funds for these research organisations are provided through a contract between the ministry and the organisations. Two of these organisations – the Centre National de la Recherche Scientifique (CNRS) and the Institut National de la Santé et de la Recherche Médicale (INSERM) – each have a significant number of research units assigned to different universities, and so part of the funding for these two institutions contributes indirectly to the financing of research in the universities in which the research units are located.

Every four years, through the Contrat d'Établissement, public funds – including research funds – are allocated to universities. In many cases, the contracts are negotiated by three parties: the MNERT, the university and the CNRS. With respect to research, by way of negotiation an agreement is sought as to how university research strategy relates to the State's research policy, and the extent of the State's commitment to providing the necessary budget and new positions to achieve its objective. In this way, the contract is expected to safeguard the coherence between university research objectives and those of the State. Negotiation is based upon the achievements of the previous contract, which are evaluated by an independent body (the CNE) as will be discussed below.

Evaluation performer and evaluation purposes

There is no evaluation mechanism specifically designed to evaluate university research. Rather, universities are evaluated by an external independent body in terms of their extensive activities, including teaching, research and management. The independent body responsible for evaluation of the education sector is the Comité National

d'Evaluation (CNE). It was established by an Act of Parliament on 26 January 1984, and later, on 10 July 1989, it was granted the status of Independent Administrative Authority with full financial autonomy by another Act of Parliament. It manages its own budgets, and logistic and management capabilities; it takes full responsibility for the methods it uses and the workplan it adopts. The committee consists of 17 members, including its president, who are appointed by government to serve for a non-renewable period of four years. Eleven of its members are representatives of the scientific community, chosen from a list of nominees from directors of various sections of the National Council of Universities, directors of various branches of the CNRS and the Institut de France. To carry out evaluations, the CNE is equipped with a permanent staff (the 'Secrétariat Général') of about 25 who are responsible for the methodological and administrative work.

The CNE evaluates not only universities, but also all institutions (schools and other education institutions) under the MNERT. The focus of the evaluation is all aspects of management and the activities of the institution. Thus, it does not specifically focus on research. Its purpose is to evaluate how effective the management teams of the institutions are in running the organisations, and in achieving their objectives. The evaluation produces 'recommendations' as to how to improve the quality of the institution. The implementation of these recommendations is the responsibility of both the institution evaluated and the Ministry. As noted before, evaluation reports also serve as information on which negotiation for a Contrat d'Établissement is based.

It should be mentioned that, in addition to the CNE, there is another institution that is specifically assigned to evaluate research activities in research units under the CNRS and other research institutions[19]. The Comité National d'Évaluation de la Recherche (CNER) was established in 1989 on the same basis as the CNE, that is, independent from the Ministry of Research and reporting annually to the President. The evaluations it carries out cover individual researchers and CNRS laboratories. In addition, it reviews trends in science and the prospects for the future development of science at the CNRS.

Evaluation criteria and methods

Here only the evaluation conducted by the CNE will be discussed as it is directly related to universities. Each particular evaluation activity is assigned to two members of the committee who then appoint outside experts to conduct a 'peer review' of the quality of

[19] See, among others Schweighofer (1997) for a detail description.

education and research and of the mechanisms and structures of management at all levels, including programmes of study, professional objectives, environment, facilities, in-service training for teachers, adult education, national and international influences, etc. For the purpose of the evaluation, the CNE, in close association with the National Conference of University Presidents, has developed a set of relevant indicators.

The university to be evaluated is required to prepare in advance a preliminary 'information package' and a set of indicators. Numerous disciplinary experts (between 10 to 15 per evaluation) are called upon to evaluate different departments. The experts' freedom of judgement is guaranteed by keeping experts' reports confidential. At the final stage of the evaluation, a plenary meeting is held where the final report is presented for approval by the full CNE. It is then submitted to the President (Vice-Chancellor) or Director of the institution, who, in turn, is given an opportunity to give his or her own comments, which are included at the end of the report. The report is then submitted to the President of the Republic and made public.

4.3 Germany

Academic research in Germany is conducted in both sectors between which the higher education system is divided. A great number of research activities are conducted by 90 'scientific universities' (wissenschaftliche Hochschulen) with a long-standing tradition of research according to the Humboldtian ideas of the mutual relationship between teaching and research. The other sector comprises 146 polytechnics (Fachhochschulen). Although polytechnics are mainly education institutions, they are undertaking an increasing amount of research. Whereas universities receive core funding for research from government, the core funding to polytechnics from government does not include any element for research. Research projects in polytechnics are largely funded by grants from industry or specific project grants from government institutions.

Three sources of public funding for research in the higher education system can be distinguished. The first is institutional funding (core funding) for teaching and research in the form of a block grant from government. It constituted almost two-thirds (60%) of total university expenditure in 1994. The regional (or Länder) governments are responsible for this budget. The core funding covers basic research infrastructure, budgetary and staff resources. The second source of research funding is in the form of a capital grant for building investment and large-scale equipment and is provided by the central government (Bund) and Länder jointly. The third source of research funding to universities is third party funds (Drittmittel), which are grants and contracts provided by

public institutions for specific research projects. The majority of these funds (37% in 1990) are allocated by the German Research Foundation (DFG), which is funded jointly by the Bund and the Länder. As in other countries, the third party funds are awarded based on a peer review of applications. The funds are granted to those applications considered to be showing scientific excellence and a high degree of social relevance.

Overall, institutional and capital funds are distributed based on the education profile that includes the number of students, graduates and scientific staff and the existing pattern of spending. Which aspect dominates varies from one Länder to another. In the allocation of the share for research a measure called the 'R&D coefficient' is used. It is based on past surveys of the working time spent on research and teaching of a sample of scientific staff. Thus, there are no research performance measures used to allocate research funds. Consequently, there has been no research evaluation related to funding allocation.

Only in recent years have a few Länder provided additional resources that are allocated on a competitive or performance-related basis. In particular, the state government of Lower Saxony has been especially active in initiating reviews of the universities. In 1998 a scientific commission was set up to assess research performance in the 12 universities. The initial focus was to be on the fields of biochemistry and history.

Finally, it should be mentioned, that research evaluation among Blue List research institutes (and others such as the Max Plank Institute) has become a common practice.

Evaluative remarks

In Germany, a research evaluation at the state level covering all universities has never been carried out. This is partly due to the fact that German universities are mainly financed at regional level, but it is also due to the attitude of the academic community towards a competitive atmosphere arising from a research evaluation involving comparison across all universities. Within the German academic community, "competition is not seen as a principle for advancing and encouraging research (and teaching) quality" (Campbell and Felderer, 1997). Although, during the 1980s and 1990s, there have been many individual studies on performance evaluation of university research in Germany (Daniel and Fisch, 1990), the studies did not have implications for funding allocation.

Despite the fact that in the late 1990s, both the Federal Ministry of Research and Technology and the Federal Ministry of Education and Science stressed that there was an increasing need for more research evaluation (Campbell and Felderer, 1997), up to the present there has been no comprehensive evaluation conducted. One of the constraints on evaluations of university research is structural. The German constitution has granted universities the privilege of a far-reaching autonomous status in conducting teaching and research, including the contents of teaching and research. As quoted by Campbell and Felderer (1997), Hufen arrives at the conclusion that with the exception of evaluations by students, systematic evaluations of teaching of university professors carried out by government are unconstitutional. Not surprisingly, the main actor of research evaluation in Germany so far has been the individual university. The Freie Universitat in Berlin, for instance, has already implemented its own evaluation mechanism, the result of which is used as a basis for the internal distribution of funding (Campbell and Felderer, 1997). However, many universities – such as those in the Verbund Norddeutscher Universitäten (VNU), including the universities of Bremen, Hamburg, Kiel, Oldenburg and Rostock – perceive the use of research evaluation for internal funding allocation as retrospective in the sense that it rewards those who have shown good performance in the past, while what is needed is evaluations that help universities to improve their performance.

The situation may change in the future. In 1998, the higher education sector has seen a foundation for major reformation through the Fourth Amendment to the Framework Act of Higher Education that was adopted by the German Bundestag (the lower house of Parliament) on 13 February 1998 (BMBF, 1998). Its aims are two-fold. First, to make competition possible through deregulation, performance orientation and the creation of incentives. Second, to secure the competitiveness of the German higher education sector in the 21st century. The amended Framework Act for Higher Education provides the legal basis for important structural improvement to higher education, the central element of which includes the introduction of higher education financing based on teaching and research achievement, as well as of comprehensive reviews and evaluation of education and research. This legal basis abolishes the ‘immunity’ of professors against any kind of external evaluation, which was guaranteed by the Act. Hence, in the near future, external university research evaluation is more likely to take place.

4.4 Italy

The higher education sector in Italy consists of 65 institutions. Of these institutions the state owns 45 universities, 3 polytechnics, 3 single-discipline universities, 2 universities for foreign students and 3 Scuole superiori (comparable to the French Écoles Normales). The remaining 9 institutions are private: 5 universities and 4 single-discipline universities (Boffo, 1996).

Government finances university research through a dual system. On the one hand, the Ministry of Universities and Scientific and Technological Research (MUSRT) provides the basic institutional funding for universities, which includes a small portion for non-directed research. On the other hand, MURST allocates funds for university research projects through national university grants and grants distributed by the Consiglio Nationale delle Ricerche (CNR) on a competitive basis. The research funding from the first part of the system is distributed to each university according to the numbers of teaching and research staff, while the funding for targeted research is allocated through peer review of applications. In addition, CNR also allocates competitive funds for targeted research that are accessible not only to universities but also other research institutions. Thus, with the exception of a peer review process for applications, so far there has been no specific evaluation mechanism of university research related to the funding allocation mechanism.

However, the National University Evaluation Council (NUEC) has recently suggested that the funding allocation should be connected to the following variables (Biggeri and Scarpitti, 1998):

1. teaching demand in each subject area;
2. results achieved in the teaching activity;
3. results achieved in the research sector.

NUEC is currently developing measures for all these criteria. In particular, measures of results in the research sectors have never been used as there are no relevant data. In order to collect this information, the council is organising a study on scientific production in the last two years for each university.

Evaluative remarks

The formal government statement of the need for university evaluation can be traced back to Act 168 of 1989 which, in addition to providing a legal basis for establishing the MURST, also included provision for the establishment of a central office for university

evaluation (Boffo, 1996). However, the National University Evaluation Council (NUEC) was established by Ministerial Decree only in 1996. The Council, consisting of five members appointed for three years, has as its main function and duties to cover all activities related to the evaluation of universities. It is an independent body, and has its own technical and administrative secretariat and its own budget, and may appoint teams of expert external bodies to carry out specific research or studies; it is more or less like the CNE in France.

It should be noted that, three years before the establishment of the NUEC, Act 537 of 1993 had already stipulated that in the future, universities would need to support any request for additional funds with evidence of results achieved. Starting from 1994, according to the Act, a growing portion of resource allocation would be based on the degree of effectiveness and efficiency of the university, both in teaching and research. Therefore, universities should provide evidence of efficiency and effectiveness through a specific evaluation exercise. However, this proviso has been largely ignored, as the practice in the following years showed that universities received approximately 8% of additional resources, regardless of results (Boffo, 1996). Thus, so far no specific attention has been given to evaluation. Nevertheless, research evaluation may become more common in the future, as an implication of the currently proposed funding arrangement by the NUEC, which will relate funding to performance.

4.5 Austria

As in many countries, public funding for universities has been very important. The annual federal budget for universities, covering almost all university expenditure, is approximately ATS 30 billion, which is much more than university income from external sources which amounts to ATS 1 billion annually. Austria is among the few countries that base the institutional operating grant allocation to universities not on the strict use of a particular formula, but on negotiation. Although the number of courses offered and the number of students enrolled are taken into consideration during such negotiations, there is no strict relation between these items and the amount of funding allocated. In addition, as in other countries, universities can pursue public research funding made available for specific targeted research and projects, the allocation of which is based on the peer review of research applications.

Decisions about the size of the annual university budget are based on a series of negotiations up the hierarchical structure of the higher education sector, beginning at the lowest level of university and ending at the top level of the Federal Ministry of Science

and Transport. Each university begins by establishing a budget request, and submits the request up the hierarchy. After the Federal Minister of Science and Transport and the Federal Minister of Finance have negotiated the amount of funding to be allocated to the university sector, the total budget of each individual university is then negotiated between the respective rectors of the universities and the Ministry of Science and Transport. Thus, it can be said that there is no research evaluation in relation to the funding allocation process.

Evaluative remarks

The state of play of university research evaluation in Austria is fairly similar to that in Germany. In Austria research evaluation is underdeveloped and there has been no attempt at systematic evaluation. There seem to be two essential barriers to university research evaluation. First, the academic community tends to resist systematic evaluations, especially those conducted by external agencies. Secondly, there has been a lack of determination at ministerial level (Steiner and Sturn, 1995).

So far, there has only been an evaluation of one particular research field: physics and electrotechnics. The evaluation in physics was by international peer review complemented by selected indicator methods, while that for electrotechnics was conducted using a less sophisticated method. Neither evaluation had any formal consequences. This was very unfortunate, especially in the case of physics where the evaluation was very thorough, involved substantial financial resources and resulted in some concrete recommendations (Steiner and Sturn, 1995).

4.6 Nordic countries: Finland, Norway, Sweden and Denmark

As in many countries, universities in the Nordic countries play an important role within the national research system. Although less than 25 per cent of total R&D expenditure is appropriated by universities, they are important given their responsibility for basic research and training scientists. Through the various countries, research in universities is financed publicly through a dual funding system including government institutional funding and various funding schemes from research councils. The allocation of research council funds is based on peer review of applications and characterised by involvement of international experts on the review panel. With regard to government institutional funding, this is distributed to universities as a lump sum, with each university having discretion as to its internal distribution. In terms of performance-based funding, in these countries, only in Finland and Denmark do the funds contain a performance-based component. However, in both these countries the performance measures are mainly

quantitative ones – for instance, the amount of external funding – that do not require a special evaluation mechanism.

In Denmark, since 1994, a new budgeting system has been in operation, by which a clear distinction is made between funds for teaching and those for research. The budget contains five elements: a basic grant, a performance-related grant for teaching, a research grant, a grant for other activities and a capital grant. Until 1995, research funds were allocated based on incremental funding. Since 1995, the amount allocated has depended on the size of the university's income from both educational activities and external research funds; no other research performance measures are adopted. The number of PhD students is adopted as a measure to determine the performance-based grant for teaching.

In Norway, universities receive block grants from government with no distinction between teaching and research. Universities are free to allocate the funds internally. Until recently, there has been no serious attempt to adopt research performance measures in the funding allocation mechanism, with the exception that universities receive a fixed amount of money per doctoral graduate. Similarly, in Sweden performance-based research funding has not been implemented. With the introduction of a new higher education law in 1993, examination results are the only performance indicator taken into account in deciding the allocation of funding.

In light of the fact that Finland represents the country with the most extensive implementation of performance-based funding, the funding mechanism and research field evaluations in Finland are discussed here in more detail. It should be mentioned, however, that research field evaluation (which is not linked to funding allocation) initiated by research councils is a common practice in the Nordic countries. These countries adopt similar methods (Luukkonen, 1995; Helander, 1995), hence the discussion in relation to Finland can be perceived as representative of the region.

Finland[20]

Finland has 20 universities – 10 multidisciplinary institutions, 6 specialist institutions and 4 art academies – all of them State-run and engaged in both education and research. The higher education system is currently being developed on two fronts, namely the university system and the network of polytechnics (AMK) that is being established. The polytechnic reform will be completed by the end of the 1990s, upgrading most

[20] This section is based on "Management by Result" (publication by the Ministry of Education) and information on the web at http://www.minedu.fi/eopm/hep

vocational higher education to the polytechnic level. The current vocational post-secondary and higher level systems will then be dismantled. The target is to have some 30 permanent polytechnics by the end of the decade. These will not engage in basic research or offer postgraduate researcher education, nor will they have professorships.

Research is conducted in universities and is financed through the Ministry of Education and four research councils under the Academy of Finland, as well as contract research with industry or government institutions. The Academy funds are allocated based on peer review of applications for research projects. Universities' appropriation of research funding from the Academy in 1997 amounted to 80% of the total distributed research funding.

General institutional funding (for teaching and research) is provided by the Ministry of Education through a contract agreement. The whole arrangement for establishing the contract between universities and the Ministry is known as 'Management by Result'. It has been developed since the late 1980s by, for instance, assessing part of the university budget on the basis of performance. Budgeting by result agreements were introduced at all universities in 1994, and were evaluated by a broad-based co-operation group from the Ministry and the universities. On the basis of this assessment, management has been improved by the adoption of a three-year agreement term, as of 1998-2000. The agreement covers the results expected to be achieved by the university and the level of the appropriations for the operating funds. The three-year agreements are revised annually by an annex agreement which has an influence on the budget for the following year.

The operating expenditure agreed upon comprises basic funding (90%), project funding (7%) and performance-related funding (3%). It is envisaged that the portion of performance-related funding will increase. Basic funding is intended to cover salaries and facilities costs. A formula-based funding model has been implemented since 1997. The connection between teaching and research is made explicit in the formula. The teaching component is represented by the target numbers of Masters degrees and that of research is represented by the target numbers of doctoral degrees. Project funding is earmarked for national programmes defined by government. Performance-related funding is paid based on agreed indicators, which focus on the quality and impact of research and education. This includes centres of excellence in research, education, artistic activities and adult education; financing awarded by the Academy of Finland; international financing and activities; graduate placements; and university-specific assessment of the achievement of the targets set and the strategic planning in the

university. All of these data are contained in the KOTA higher education database, based on the report submitted by universities each year. Thus, no specific evaluation system is required to determine the funding allocation.

Evaluation performer and evaluation purpose

To assist universities and polytechnics in matters related to evaluation, the Finland Higher Education Evaluation Council (FINHEEC) was established in 1995. Evaluations of the higher education sector are usually classified into three categories: institutional evaluation, programme and thematic evaluation, and accreditation. None of these evaluations is targeted specifically at university research. The institutional evaluation does not use a uniform model for all the universities. In this way, the government recognises the differences between universities and emphasises the developmental role of the evaluation. Nevertheless, most of the evaluations can be characterised as broad institutional evaluations of the basic preconditions for teaching and research, and the capacity for change. These preconditions include a statement of the institution's mission and goals, key processes, institutional arrangements, and resources and performance, with less emphasis being given to the last two than to others. The emphasis in evaluation also varies from one university to another: one university might lay special emphasis on the evaluation of teaching; while another might emphasise the strategies, regional role or administration.

The body that has been actively evaluating research activities in Finland is the Academy of Finland, which in the early 1970s started to carry out research evaluations. However, it does not specifically evaluate university research; most evaluations have been initiated outside universities. It might be said that until recently no systematic and nationwide ex-post evaluation of university research has been implemented in Finland (Felderer and Campbell, 1998). Traditionally, evaluation has focused on individual scientists, research projects and teams. More recently, evaluation on research fields that include research groups within universities was conducted. Up to the present, 18 scientific research fields have been evaluated. The first evaluation on inorganic chemistry was conducted in 1983, followed by (to mention just a few examples) experimental nuclear and high energy physics (1985), automation technology (1986), peace research (1990), research on climate change (1992), legal science (1994), and research on molecular biology and biotechnology (1997). The evaluations of research fields have emphasised the assessment of research excellence in terms of international prestige and contributions to the forefront of science. Such evaluations are driven more by the need to improve the quality of science, than by pressures to cut the budget.

In 1993, a new policy to nominate centres of excellence in research was introduced. As noted earlier, being a centre of excellence is one of the criteria for the performance-related funding component. The centre of excellence may consist of research groups, research centres or a larger umbrella organisation and networks. The Academy of Finland, FINHEEC and the Ministry of Education have been responsible for the selection process, which of course involves some form of evaluation. As of 1997, 17 research units and departments had been nominated and had received extra funds.

Evaluation criteria and methods

The evaluation conducted by FINHEEC generally consists of three phases: first, the university carries out a self-evaluation, which results in a report. In the second phase, the self-evaluation report is assessed by an external evaluation team and followed by a visit to the university. The third phase constitutes the production of a final report by the external team.

To illustrate the method adopted in research field evaluations by the Academy, the evaluation of the 'electronic research' field will be described based on the report by Felderer and Campbell (1998). In September 1995, the Academy of Finland commissioned the Research Council of Natural Science and Engineering to evaluate Finnish electronic research. Following this, a committee was set up, which, in its turn, appointed two international experts to lead the evaluation. For the purpose of evaluation, the scope of electronic research was limited to pre-defined sub-areas on which the evaluation concentrated. Based on this division, 28 research groups from universities and research institutions were identified and subjected to evaluation. The research groups were evaluated with respect to the following six aspects:

1. The mission, vision and goals;
2. The supply of resources and the efficiency of their use;
3. The scientific competence and the degree of innovation;
4. The technological competence and the co-operative activities with other research groups, industry and users of research results;
5. The national and international importance of the research group and of their research results for the scientific community and for the further qualification of researchers;
6. The relevance of the research group and their research results for industry.

Methodologically, like the institutional evaluation carried out by FINHEEC, the evaluation procedure consisted of three phases. First, a questionnaire was distributed to research groups. It was designed to identify comprehensively the performance and the

use of resources by the research groups. Having examined the result of the questionnaire in the second phase, the evaluators conducted a visit to each research group and interviewed members of the research groups. In the third phase, based on all this written and oral information, the evaluators summarised their findings in a report written in English. The research groups were given an opportunity to comment on the report before it was finalised and published. In addition to specific recommendations for individual research groups, the report also analysed the current state of electronic research in Finland.

With regard to the nomination of a centre of excellence, an international peer review method is adopted. The main criteria used in the selection are:

1. The national and international position of researchers;
2. The scientific significance, innovativeness and effectiveness of research;
3. The quality, quantity and focus of scientific production;
4. Patents;
5. The national and international mobility of researchers; and
6. The number and level of foreign researchers in the centres.

Recently, the different nature of various disciplines has been taken into account explicitly, and the above general criteria have been adjusted to the specific nature of the discipline being evaluated.

Evaluative remarks

It is interesting to note that as a possible future development of the existing funding system, a performance-based funding mechanism similar to the RAE was suggested by the Ministry of Education. The suggestion was contained in the report 'Management by Results' of the working group assigned by the Ministry to outline the future trends and the need for developing the existing system. The working group suggested that 35% of operational funding should be allocated on the basis of the performance of research. The research performance of research groups within all universities would be evaluated by the Academy of Finland every three years using the peer review method. Following the British model, based on the evaluation results, the research units would be given a five-point score from one (lowest) to five (highest), which, in turn, would determine the amount of funding allocated to the units (Kaukonen, 1997).

The proposed evaluation mechanism was heavily criticised by almost all universities and relevant parties, which resulted in the suggestion being frozen by the

Ministry. The main objection was that the mechanism would give the Academy of Finland an overwhelming influence over policy, which was perceived as problematic. Furthermore, it was argued that the proposed system would result in a normative standard for resource allocation among universities (Kaukonen, 1997).

4.7 Eastern European Countries: Hungary, Poland and the Slovak Republic

Throughout Eastern Europe, after World War II, research systems can be characterised as being based on the Soviet tripartite system (Frankel and Cave, 1997). Within this system there is a clear division between institutions conducting teaching, basic research and applied research. Universities focused on teaching activities, basic research was conducted in institutes of the Academy of Sciences, and applied research was conducted in institutes under industrial ministries and other government agencies. Governments in the region also adopted the Soviet system of funding research largely through block grants to research institutes, where the scientist-administrator had great power over how the funds were distributed to individual research teams. In such a system, favouritism and political connections gave rise to the funds being distributed to poor-quality research teams with the right connections, while high-quality research teams could often be under-funded (Frankel and Cave, 1997).

This research and funding system has been changing radically since the early 1990s. Since the end of the cold war, political and economic systems in Eastern European countries have been in a state of transformation from centrally-planned, non-competitive economies to more open, competitive market economies. This transformation has had a significant impact on science policy in the region. The autonomy of science in term of evaluating itself through peer review, which was completely subordinated to the central plan during the communist regime, has been restored. In addition, in the first period of the transition, the economic crisis has seen research evaluation emerge as an important tool, mainly to examine where to cut research budgets without completely destroying research activities. In the subsequent development, research evaluation through peer review has become the main evaluation mechanism used to allocate funds, leaving little room for old-fashioned funding allocation based on political connections. This method of evaluation has characterised almost all research evaluation activities throughout the region (Hangos, 1997; Zilahy and Lang, 1997).

As noted above, during the communist era, research activities were mainly conducted in institutes of the Academy of Sciences, while universities concentrated on teaching activities. It is not surprising that, in the post-communist era, although the

system has experienced major changes, the Academy of Sciences still plays a major role in research activities. Consequently, research evaluation has been mainly carried out to evaluate research at institutes of the Academy of Sciences. There has been no comprehensive evaluation of university research. In the Slovak Republic, however, there has been an evaluation mechanism in place to evaluate the research activity of departments within universities, the results of which influence the allocation of core funding (personal communication with Štefan Zajac). In the following, the evaluation of institutes of the Academy of Sciences in Hungary and the Slovak Republic will be discussed briefly, followed by a description of the funding allocation mechanism in Poland.

In Hungary, for several decades the Hungarian Academy of Sciences has conducted a fairly comprehensive evaluation of all its institutes almost every year. However, the evaluation in 1992 had an important impact as it coincided with the cutting of research funding as a result of the economic crisis. The evaluation was conducted in two phases. In the first phase, each individual research institute was evaluated with the aim of providing a basis for restructuring the Academy's research network and reformulating its tasks. In the second phase, each individual organisational research unit and research group within the institutes was evaluated. This provided the basis for a more differentiated distribution of funds. The method adopted in the evaluation was peer review combined with quantitative indicators. The main findings of the evaluation led to recommendations with regard to several issues: the main task of the Academy's research network, management of human resources, financial conditions and organisational changes (Zilahy and Láng, 1997).

In the Slovak Republic, in 1992, both the Academy of Sciences and the universities set up an accreditation committee to evaluate research activity in research institutes and university departments. In the evaluation of Academy institutes, the following quantitative indicators were used as the main criteria (Tino, 1997):

- a list of scientific publications from the previous five years, classified into monographs, original research, and survey articles; in addition, institutes were asked to select ten representative publications;
- the number of citations (included in the SCI) during the previous five years;
- participation in journal editorial boards;
- participation in conferences (domestic and foreign), conferences organised by institutes, membership of international scientific organisations, and relations with the international science community (agreements on co-operation, joint projects etc.)

Based on these quantitative data, research institutes were evaluated and classified into several groups according to their level of scientific activity.

For the evaluation of their departments, universities employed the indicators used by the Academy of Sciences, with the following additions:

- titles and scientific qualifications of all teaching staff;
- number of postgraduate students;
- research activity of senior research staff;
- post-university employment of graduates and number of applications from foreign students.

In addition to criteria related to research, faculties are also evaluated in relation to teaching activities and managerial aspects, such as financial discipline. Since 1992, every three years this evaluation has been conducted under the supervision of the Accreditation Committee by universities. Faculties are classified into four groups: A (best), B, C and D (worst). Each group is assigned a weighting that determines the amount of money it receives. The allocation of funding is based on the student numbers, adjusted to take into account the grouping. Assuming that their student numbers are the same, universities in group A will receive more money than other groups (personal communication with Štefan Zajac).

Poland

In Poland,[21] after the fall of communist regime in 1989, the new system for managing and financing Polish scientific research was based on the Act establishing the Committee for Scientific Research (CSR) of 12 January 1991. The chair of CSR is appointed by the parliament, and two-thirds of its members are active researchers elected by the scientific community; the rest are members of the Council of Ministers appointed by the Prime Minister. This new institution is responsible for formulating and executing science policy, including funding allocation. All state budgetary scientific funds are administered by the CSR. It allocates research funding through several competitive channels. All institutions, including faculties in universities, have to compete for funds.

[21] The discussion on Polish funding mechanisms is based on Jablecka (1995) and personal communication with Julita Jablecka.

Evaluation criteria and methods

Research funding administered by CSR has been an important source for universities. In addition to the general university funds (operating funds) which are allocated based on education profile, such as number of students, universities also compete on an equal footing with other research institutions for research funds through two schemes administered by CSR. First, there is a grant system for individuals and research teams that is based on open competition: everyone who has completed postgraduate education (at least to the MSc level) is eligible to apply for a research grant. The applications are reviewed by panels consisting of active researchers; as such, it is traditional peer review.

The second scheme is the so-called statutory funding that is distributed to faculties within universities (e.g. mathematics, languages, sociology). This funding is based on an ex-post evaluation exercise. Annually, institutions submit applications consisting of the record of the past year's achievements and the research plan for the coming years. The assessment is conducted by expert panels (Subcommittees) that include only scientists. As a result of the evaluation, institutions are assigned to a category (A for the best institutions, C for the poorest institutions). The final funding allocation is decided by the Committee for Basic and Applied Research, which is a committee under the CSR. The relative proportion of funds received by units of classes A, B and C is, respectively, 1.151, 0.951 and 0.550.[22] It should be noted that the other research units (institutes of the Polish Academy of Sciences and government establishment) are financed according to the same statutory formula funding and the same evaluation criteria and indicators as faculties of universities.

In 1998 for the first time a new formula was proposed and introduced for statutory funding. Until 1998 the funding level for evaluated units was determined according to categorisation of institutions (the so-called algorithm) based on a combination of quantitative and qualitative methods (the latter were criticised for their subjectivity). A new 'parametric system' is based almost exclusively on quantitative methods. It consists of the total of the number of points for **performance R_p** and for so-called **general results R_g.**

[22] Committee for Scientific Research (KBN), http://www.kbn.gov.pl/en/general/reseval.html

R_p consists of 6 groups of indicators. Points are awarded for:

1. the number of publications in refereed journals;
2. publication of books (monographs);
3. scientific degrees awarded to academic personnel in the unit;
4. number of patents;
5. implementation of research results; and
6. a right (licence) to carry out quality evaluation or accreditation of national laboratories.

Within each group of performance indicators, various kinds of indicators are distinguished: for instance, from a group of publications there are papers published in journals on the list of the Science Citation Index (SCI) of the Institute for Scientific Information in Philadelphia, papers in other refereed journals of institutional scope, publications in Polish journals, and so on.

The following indicators are taken into account for the general result indicator R_g:

- various research projects (grants);
- research commissioned at the unit;
- research projects financed from abroad;
- international co-operation agreements;
- numbers of long-term scientific visitors from abroad;
- numbers of citations;
- awards for scientific or practical achievements;
- expert reports commissioned from the unit;
- the right to award academic degrees by the unit;
- dissemination of knowledge among the lay people (e.g. presentations in popular journals);
- existence of doctoral studies organised in the unit;
- organisation of international and national conferences.

The total number of points for performance indicators, $R = R_p + R_g$ is divided by the number of staff employed (N) – in such a way that the indicator of effectiveness (E) is determined. This is the basis for determining, every three years, the category of institutions (5 categories) and subsequently the level of funding. It should be noted,

however, that the new algorithm of statutory funding is now under discussion (including the weights ascribed to each of the indicators) and will probably be changed.

5. University Research Evaluation and Funding in North America

Among the North American countries, the United States is the only one with an evaluation programme at the state level. In Canada and Mexico there is no university research evaluation conducted at the state level. In Canada, universities are governed by provincial governments. At the federal level, a law has recently been approved that requires federal departments (including research centres) to produce performance reports. This law does not apply to universities. So far, the Association of Colleges and Universities engages only in collecting a lot of statistical data but not for the purpose of evaluation. Similarly, until recently, no provincial government has carried out a university research assessment. The first Canadian research assessment will be conducted in summer 1999 by the Office of Science and Technology (OST) and will involve 14 universities (personal communication with Benoit Godin). However, the data will be owned by the universities and will not be used for the purpose of resource allocation. In the case of Mexico, evaluation is mainly conducted by each individual university. Each faculty has its own Academic Committee or Technical Council that is responsible, among other things, for evaluation. A recent attempt has been made to evaluate the research performance of a number of research units within the National University of Mexico (UNAM) (Lopez-Martinez and Rocha-Lackiz, 1998). In the following we briefly discuss research evaluation and funding in the USA and Canada.

5.1 The United States

The United States higher education sector consists of well over 3,600 universities and colleges. However, research is concentrated in a much smaller number of universities and institutes of technology or polytechnics, usually those that offer graduate degree courses and, in particular, those awarding doctoral degrees. In 1995, 200 institutions accounted for 94% of all university R&D expenditure (Senker, 1999). Unlike many countries in Europe where the dual funding system for university R&D is widely adopted, there is no formal dual funding system in the United States. The institutional funds that public universities receive from state government are mainly for teaching, not research.

The federal government provides the majority of research funding for universities through various federal agencies on a competitive basis. It was estimated that in 1997 the federal support for university research amounted to $14.3 billion, which was about 60% of the total. The other sources of research funds were academic institutions' own funds $4.5 billion (19%), state and local government $1.8 billion (7.5%), non-profit organisations $1.8 billion (7.5%) and industry $1.7 billion (7%). Of

the total federal funds, 82% were provided by three federal agencies: the National Institute of Health (NIH) 52%, the National Science Foundation (NSF) 15%, and the Department of Defence (DOD) 10%. Small amounts were also provided by three other federal institutions: the National Aeronautics and Space Administration (NASA), the Department of Energy (DOE) and the Department of Agriculture (USDA) (Senker, 1999).

Federal funds for university research are allocated based on the review of applications, which might request funding support for an individual, a specific project or a centre. The review mechanism can range from 'peer review' by an expert in the related field, to 'manager discretion', relying on decisions by management in the funding body.

Research-Doctorate Programmes Assessment:[23] ***Evaluation performer and purpose***

As research funds are allocated based on the review of applications, there is no university research evaluation conducted that is related to research funding allocation. However, research evaluations in various forms have been widely conducted by a number of funding agencies. Evaluation activities might take the form of peer review, research project evaluation or programme evaluation, with the main objective being to evaluate the progress of funded projects. These evaluations are not specifically designed for university research, although of course some research groups within universities are involved in the evaluations. Thus, with regard to university research, so far there has been no evaluation covering all universities in the United States.

However, since the 1960s, research-doctorate programmes within universities across the country have been the subjects of a national study, which includes an assessment of the quality-related characteristics of the programmes. The most comprehensive and comparable studies were conducted in 1982 and 1991-94. The two studies of research-doctorate programmes were commissioned by the Conference Board of Associated Research Councils, which consists of the chief executive officer and one representative from each of the American Council on Education (ACE), the American Council of Learned Societies (ACLS), the Social Science Research Council (SSRC), and the National Research Council (NRC).

The 1982 study "*An Assessment of Research-Doctorate Programs in the United States*" generated various statistics about the characteristics of 2,699 research-doctorate programmes in 32 fields of study, including peer reviews of programme quality. The

[23] This section is based on Goldberger *et. al*. (1995).

statistics proved to be an important source of information for educators and policy makers. In the early 1990s, it was thought that a new study was needed as the enterprise of doctoral education had continued to grow since 1982. The number of institutions providing doctoral programme increased from 325 in 1980 to 364 in 1992, and many institutions have expanded the number of their doctoral programmes. All these changes raised the need for an update. In 1990, the Conference Board of Associated Research Councils asked the National Research Council (NRC) to undertake a study on the research-doctorate programmes that would update and expand the 1982 study. After a planning phase, in 1991 the council appointed the Committee for the Study of Research-Doctorate Programs in the United States, consisting of 16 members, to undertake a three-year study, taking the 1982 assessment as a starting point.

The 1991 study sought to provide descriptive statistics and other quantitative indicators of research doctorate programmes in four areas:

1. The students participating in graduate programmes – their demographic characteristics and career outcomes;
2. The faculty that provide the educational and research leadership in research-doctorate programmes;
3. Institutional resources brought to bear in creating an enriched environment for preparing research scholars and scientist; and
4. The views of disciplinary peers regarding the effectiveness of those programmes in preparing students for careers as research scholars/scientists.

Thus, the main purpose of the study was to produce statistics on the characteristics of research-doctorate programmes, which would be of use to a wide range of parties interested in the programmes. The parties include policy makers, students, educators, administrators and also higher education researchers. Therefore, the data had to be presented in such a way that it was accessible to all those parties. Also included in the study was a survey of faculty members who were asked to give their judgement about the quality of research-doctorate programmes. Each faculty member was asked to rate 50 programmes in his/her field with respect to two dimensions of quality: ‘scholarly quality of programme quality’ and ‘effectiveness in educating research scholar/scientist’. Based on these, the programmes were then ranked nationally. In addition, the overall score and rank of each university was also published. This ranking is not related to the funding allocation to universities.

Evaluation criteria and methods

The study used two methods to assess programmes: survey work and review of available descriptive statistics. The committee selected the fields to be include in the study based on a combination of three conditions: the number of PhDs graduating nationally, the number of programmes training PhDs within a field, and the average number of PhDs per programme. In addition, the committee adopted the criterion of robustness, that is, to be included in the study, a field had to have awarded a minimum of about 500 degrees in about 50 programmes for the years 1986 to 1990. As a result of these conditions, 41 fields were included, grouped under the following categories: Arts and Humanities (11), Biological Sciences (7), Engineering (8), Physical Sciences and Mathematics (8) and Social and Behavioural Sciences (7).

Having determined the fields, the committee identified eligible institutions, which were any institutions within a field that had produced at least three PhDs between 1988 and 1990 and one in 1991 or that had received a rating of 2.0 or better in that field in the 1982 assessment. As a result, 300 institutions were identified. The committee then wrote to the presidents of the eligible institutions inviting them to participate in the study. In addition, the presidents were asked to appoint one person from their institution to serve as the Institutional Co-ordinator (IC), with whom the committee could work during the subsequent stages of the study. Sixteen universities either did not respond to the invitation or declined to participate.

Institutional co-ordinators at 284 institutions were sent a list of the programmes at their institutions that were eligible for participation, and a questionnaire. The IC was also given a chance to nominate additional programmes in the event that the committee had overlooked strong programmes in their institutions. Ten institutions failed to provide information within the time-frame established by the committee. This process of selection resulted in the participation in the study of a total of 3,634 programmes in 41 fields at 274 universities. This represents about 78,000 faculty members providing training through these programmes to 90% of the total number of PhDs produced in these fields between 1986 and 1992.

The first step in the assessment was the General Survey of Graduate Faculty, which was conducted in the spring of 1993. Survey forms were sent to a sample of faculty members, chosen from lists provided by the ICs. Approximately 19 per cent of faculty members included in the study were sent a questionnaire; the return rate was about 50%. Each faculty member was asked to rate on a six-point scale (0 = not sufficient for doctoral education; 5 = distinguished) the quality of 50 programmes in

his/her field, with respect to two dimensions of quality: 'scholarly quality of the programme' and 'effectiveness of the programme in educating research scholars/scientists'. From these responses the committee calculated the mean rating for each programme.[24]

The second part consisted of identifying and collecting data describing the key features of each of the 3,634 participating programmes. In this effort, attention was focused on variables thought to be related to the quality and effectiveness of the programme, such as publications, citations, the number of PhDs produced, career outcomes of graduates, numbers of enrolled students, etc.

Evaluative remarks

As with any other evaluation activity, the results of the research-doctorate programmes assessment did not escape criticism.[25] Notwithstanding, the study led to the following main conclusions:

- With respect to the relation between the size of the programme and the rating, the study indicated that the top-rated programmes in most fields tend to have a larger number of faculty and more graduate students than the lower-rated programmes.

- The vast majority of research-doctorate programmes included in the study had faculty who received some type of federal support for research between 1986 and 1992.

- Most doctoral programmes in the Sciences and Engineering had faculty actively engaged in the publication of scholarly work. However, there was strong evidence of a strong positive correlation between the 'scholarly quality of programme faculty' and the citations of their work. Faculty in top-rated programmes are cited more often than faculty in low-rated programmes.

[24] The mean rating for scholarly quality are as follows: (less than 1) = not sufficient for doctoral education; (1.00-1.99) = marginal; (2.00-2.50) = adequate; (2.51-3.00) = good; (3.01-4.00) = strong; (4.01 +) = distinguished. For effectiveness, the mean ratings are as follows: 0.0-1.49) = not effective; (1.5-2.49) = minimally effective; (2.5-3.49) = reasonably effective; (3.5-5.0) = extremely effective.

[25] See among others Ehrenberg & Hurst (1996), Maher (1996) and Webster (1996).

- A larger share of faculty in the top-rated programmes in the Arts and Humanities are likely to have received prestigious awards or honours than faculty in the lower-rated programmes (this pattern is not the same for other fields).

5.2 Canada

In 1993-94, the higher education sector in Canada consisted of 77 universities and 206 colleges spread unevenly across ten provinces and two territories. The Province of Ontario with 21 universities is the province with the largest number of universities, followed by Nova Scotia and Quebec with 13 and 8 universities, respectively. The provincial government has an exclusive jurisdiction to govern education in the province, including universities and colleges. Each province has its own ministry of education or a ministry responsible for education. Accordingly, the main source of funding for universities is the provincial government. Provincial government funds for the core operation of universities are allocated through a block funding formula, based mainly on the number of students enrolled. This core funding covers the cost of teaching activities and research infrastructure and equipment.

On the other hand, the federal government provides funds for targeted research through three federal research councils: the Natural Sciences and Engineering Research Council (NSERC), the Medical Research Council (MRC) and the Social Sciences and Humanities Research Council (SSHRC). The allocations of these funds are based mainly on peer review of applications. They constitute a principal source of university research funding. Other sources of federal research funding for universities include various federal government departments through sponsored research conducted under contract with the departments.

For the various mechanisms by which these various funds are allocated to universities – block formula funding based on the numbers of students, peer review and contracts – there is no research evaluation mechanism related to their allocation. Moreover, although there are clear gradations in both research capability and levels of research funding among universities within each provincial system, most suggestions to formalise these distinctions have been greeted with alarm by the research community (Wolfe, 1998). This might indicate an opposition among the university research community towards any attempts to evaluate their performance.

Evaluative remarks

Quality of education has been the concern of provincial government. However, research evaluation has been conducted neither at the provincial nor the federal level. The Canadian Association of Colleges and Universities (AUCC) collects a lot of statistics on universities, but does not use them for evaluation. The first federal research assessment is being conducted in the summer of 1999 by the Office of Science and Technology (OST) for 14 universities. However, this evaluation will not be related to funding allocation nor will it have any consequences for universities. The evaluation results will be privately owned by the individual universities involved in the evaluation (personal communication with Benoit Godin).

It should be noted that, as in many other countries, funding councils have carried out programme evaluations in an attempt to assess whether objectives are being realised, and to assess what impact a particular research programme is having upon the advancement of knowledge or upon society in general. Hanson (1994) suggests, for example, that while peer review is central to the allocation of research grants within a programme, programme evaluation at the SSHRC plays a fundamental role in providing feedback on the continuing relevance, outcomes and alternatives to the funding programmes themselves.

6. University Research Evaluation in the Asia Pacific Region

In the Asia Pacific region, Australia and Hong Kong are the only countries that carry out a research evaluation nationally and use the result to decide on the allocation of funding (Bourke, 1997; Atkinson and Massy, 1996). However, while the method adopted in Hong Kong to a certain extent is similar to the RAE in the UK, it should be pointed out that in Australia the evaluation is concerned mainly with the quantity of research, as its name the 'Research Quantum Publication Collection (RQPC)' might suggest. With regard to the methods used for evaluation in the above countries (with the exception of Australia which uses publication counts), expert peer review constitutes the main method of assessment supplemented by various bibliometric measures.

6.1 Australia

Through the adoption of the Unified National System in the higher education sector in 1988 the number of universities eligible for government research funding has increased. The federal Australian Commonwealth Government provides the majority of university research funding through a dual system.[26] Funds are allocated through an institutional operating grant (core funding) and through a separate targeted granting scheme. Institutional operating grants are allocated to universities as block grants and universities have discretion to distribute these funds within their own institutions. They can use them for their own grants or award schemes, research equipment, infrastructure, or whatever other research activities they choose. The targeted granting scheme – Commonwealth Competitive Grant – is managed by various research councils and other government agencies, among which the Australian Research Council (ARC) is the biggest provider. There are some 40 competitive granting schemes.

In addition to research core funding, which might be spent on research infrastructure, there are two other schemes designed specifically for research infrastructure: the Research Infrastructure Equipment and Facilities (RIEF) scheme and the Research Infrastructure Block Grant (RIBG) scheme. The allocation of funding from the RIEF scheme is designed specifically to encourage collaborative research between institutions and is based on applications submitted by universities to the Committee for International and National Co-operation (CINC) for consideration and then submitted to the Minister for approval. The grants allocation of the RIGB scheme to universities is formula-based, with allocations reflecting the relative success of each university in

[26] In 1992, 91% of R&D conducted in universities was funded by the Commonwealth Government. Other funding providers are state and local Government, industry, non-profit organisation and overseas providers with funding shares of 2%, 2%, 4% and 1% respectively (Industry Commission, 1995).

gaining competitive research funds, as calculated by the National Competitive Grants Index (NCGI) (data collected by DETYA through a particular procedure as will be described below).

The objectives of the infrastructure funding schemes are fourfold:

1. to meet project-related infrastructure costs associated with National Competitive Grants;
2. to remedy deficiencies in the current research infrastructure;
3. to enhance support for areas of research strength; and
4. to ensure that areas of recognised research potential, in which institutions have taken steps to initiate high quality research activity, have access to the support necessary for development.

Whereas peer review of grant applications is the main review method for allocation decisions of targeted grant schemes (Industry Commission, 1995), the core funding is distributed among universities through a formula funding mechanism. Following consultation with the higher education sector, the Relative Funding Model was introduced in 1990. In this model, funding for teaching and research training is clearly separated from research funding. The funding component for teaching activities and research training is essentially based on the number of students at each university. The research part of this core funding, which has become known as the Research Quantum,[27] was initially based on the success of universities in gaining research funding from Commonwealth Competitive Grants. However, it was later recognised that success in obtaining funding from the Commonwealth granting scheme did not fully represent research performance, and hence the criteria needed to be broadened to include other sources of funding. In addition, measures other than funding, such as publications and higher degree completion rates, were incorporated in the formula.

Both research input measures and output measures are incorporated in the index, which is called the Composite Index. In 1993, the Minister for Employment, Education and Training announced that from 1995 the Research Quantum would be allocated on the basis of a new composite research index. This Composite Index was developed by a working group with inputs from the Australian Vice-Chancellor's Committee (AVCC), the Australian Research Council (ARC), the Higher Education Council (HEC) and the

[27] The Research Quantum is used to support research separate from those activities directly linked to teaching and research training.

then Department of Employment, Education, Training and Youth Affairs (DEETYA). Since the restructuring of Commonwealth Government Departments in October 1998, the new Department of Education, Training and Youth Affairs (DETYA) has taken over responsibility with respect to data collection and the Research Quantum calculation. It has the responsibility for managing data for the Composite Index, for calculating funding allocation and for advising the Minister on the appropriate components in the Index and the appropriate weighting of elements within the Index.

The measures contained in the Composite Index are:

1. research input measures (funding):
   - the amount of each university's funding from Commonwealth competitive grants;
   - other public sector research funding;
   - industry and other research funding.

2. research output measures:
   - numbers of research and scholarly publications produced by staff and students;
   - numbers of higher degrees completed (Masters and PhD).

These various components of the Index have been weighted differently from year to year, as the balance between them is being progressively refined. For the 1999 Research Quantum Allocation, the weightings between the categories within each component are as follows (DETYA, 1998):

Funding Source		*Weighting*
Category 1 - National Competitive Research Grants		
(i)	Commonwealth schemes (including a share of DISR funding to CRCs[28])	2
(ii)	Non-Commonwealth Schemes	2
Category 2 - Other Public Sector Research Funding		
(i)	Local Government (competitive and non-competitive)	1
(ii)	State Government (competitive and non-competitive)	1
(iii)	Commonwealth Government (other than those listed above)	1
Category 3 - Industry and Other Research Funding		
(i)	Australian	
	- Contracts	1
	- Grants	1
	- Donations, Bequests and Foundations	1
	- Syndicated Research Development	1
(ii)	International Funding	1

*Publication Category**		*Weighting*
A1	Authored Book – research	5
B	Book Chapters	1
C1	Article in Scholarly Journal	1
E1	Conference Publication – full written paper, refereed proceedings	1

*****The value of joint publication is shared equally between the authors

Degree Completion Category	*Weighting*
Doctoral degree by research	3
Master degree by research	1

The weighting scheme suggests that, in the case of funding sources, the total from funding sources category 1 is multiplied by 2 before it is added to funding from categories 2 and 3. Furthermore, for the three components (funding, publications and degree completion) different weightings are attached: 80% for funding measures, 10%

[28] A number of universities conduct co-operative research projects under Co-operative Research Centres (CRCs), which are funded by the Department of Industry, Science and Resources (DISR).

for publications and 10% for degree completion. The calculation of the Composite Index for a particular university proceeds as follows.

For 1999, the total Research Quantum funding available is $221,158,000. Supposing that University X's shares of total national funding, publication and degree completion, averaged over the two most recent years, are 4.5%, 3.65% and 5.3% respectively then the total Composite Index for university X is obtained by first multiplying the share by the corresponding weighting for each component (80%, 10% and 10% respectively), and then adding them together. The Composite Index of University X represents its share of the total universities' research activities. Hence, the Research Quantum allocation for University X is the Composite Index multiplied by the total Research Quantum available. The following indicates these calculations.

Funding share	4.50%	multiplied by	80%	gives	3.600%
Publications share	3.65%	multiplied by	10%	gives	0.365%
Degree Completions share	5.34%	multiplied by	10%	gives	0.534%.

Overall Composite Index for University X (add the above)	4.499%
Total Research Quantum available in 1999	$221,158,000
Research Quantum allocation for University X (4.449% of $221,158,000)	$9,839,319

The fairness and accuracy of the Research Quantum funding allocation mechanism relies heavily on the accuracy and quality of data collected. DETYA has developed processes to ensure the quality of the data. Each year, universities are required to submit the necessary data in accordance with procedure set up by DETYA: "Higher Education Financial and Publications Research Data Collection". The financial and publications data are collected through this procedure, whereas details of higher degree student completions are collected separately, through the DETYA Student Collection. To ensure the robustness of data, audit procedures are established for both university financial data and publication data.[29] Moreover, submission of data by universities must be accompanied with certification from the Vice-Chancellor, indicating that the data have been collected in accordance with the established procedure.

[29] These measures represent a quality control of publication data. Apparently, this quality control has been employed following some anomalies in the numbers of publications that certain universities were claiming as publications.

Evaluative remarks

In the report of the Industry Commission (1995) it was suggested that as a research funding mechanism, the Research Quantum has a number of desirable features:

- it rewards strongly performing institutions; and
- it does not require the costs of application and evaluation that are associated with other selective institutional grants.

In addition, the fact that universities maintain their discretion in distributing the funds has been highly valued by universities as a way of maintaining their autonomy.

However, the suggestion that the Research Quantum rewards the performing institutions should be examined cautiously. There is a danger in using research grant incomes – on which the Research Quantum is mainly based – as a performance measure (Gillett, 1989; Hare and Wyatt, 1988). In the first place, as is commonly recognised, research grant income is not a measure of performance, but of research input. Furthermore, Gillett (1989) asserts that the use of research grant income as a performance measure has "intrinsically low validity, and is strongly confounded with a variety of extraneous factors that are unrelated to research performance". In addition, a measure relating to the amount of dollars received, as is currently used in the Research Quantum, can be problematic and misleading across and within disciplines. It might be the case that a particular research project requires expensive equipment, which would boost its monetary value, while other more significant research might require smaller amounts of money. Thus, in some cases, measures related to the number of research grants may be more appropriate than the number of dollars received (Print and Hattie, 1997).

Compared to the RAE in the UK, which is mainly based on informed peer review, the Research Quantum funding mechanism is more mechanistic in nature. The research performance is evaluated solely on the basis of quantitative measures; no qualitative measures are involved in the evaluation. In effect, the Research Quantum is no more than a funding allocation mechanism based on the volume of research activities within universities. It is "confined to counting annually the gross number of undifferentiated entrants each institution can place in the classification categories which are weighted for funding purposes" (Bourke, 1997, p. 25). The idea to incorporate more qualitative measures was put forward by a standing committee of the Australian Vice-Chancellors' Committee in 1994. However, this required considerable developmental work, which delayed its realisation until now. In addition, as funding generated through

Research Quantum funding system has assumed greater importance to universities, any strong adjustment to the system needs careful justification (Bourke, 1997).

Bourke (1997) suggests that the RAE can be perceived as fulfilling three functions: a competitive source of discretionary income in which research is the chosen indicator because it is 'measurable'; a reward for the quality and/or volume of research output: and an instrument of research policy. The Research Quantum operates primarily as the first of these functions. It functions as a reward based on the volume of research output but it does not reward quality and is seldom used as an instrument of research policy – it is rarely used as a mechanism for internal distribution of discretionary funds. Instead, Bourke proposes a funding mechanism that combines the current practice of Research Quantum with a 'professional judgement' element of the RAE.

It should be noted that at the time of writing this report, the higher education sector in Australia has been long expecting the new government White Paper on research and research training that is expected to change current arrangements significantly. Discussion on this matter has been going on for sometime. The White Paper was supposed to be published at the end of 1998, but to date, the paper has not yet appeared (personal communication with Satis Arnold).

6.2 New Zealand[30]

There are 39 institutions within the higher education sector in New Zealand. These consist of 7 universities, 25 polytechnics, 4 colleges of education, and 3 wananga. The latter is a special education institution devoted to maintaining and developing knowledge relating to the Maori tradition. Funding for universities is provided by the Ministry of Education. About 8% of government funding to universities (which in 1992-93 amounted to NZ$108 million) represents research core funding. Other sources of research funding are the Public Good Science Fund (PGSF) and the Marsden Fund, which are provided respectively by the Foundation for Research, Science and Technology (FRST) and the Royal Society of New Zealand.

The PGSF funds are allocated through a competitive bidding system and are open to applications from any public or private enterprise or organisation, including government departments, and individuals based in New Zealand who have an ability to

[30] This section is mainly based on the White Paper produced by the New Zealand Ministry of Education "Tertiary Education in New Zealand, Policy Direction for the 21st Century" (1998).

conduct research. The fund, which in 1998-99 amounted to NZ$290.7 million, is targeted at 17 priority areas of research determined by government. The Marsden funds are also allocated on a competitive basis through peer review, and are targeted at 'Blue Sky' research. The applications are assessed primarily on three criteria: scientific merit of the proposal, the track record of the researcher that reflects the potential contribution to the advancement of knowledge, and contribution to development or broadening of research skills in New Zealand. Whereas the PGSF and the Marsden funds are allocated based on peer review of applications, the research core funding from Vote Education is currently distributed to universities as tuition subsidies based on the number of students.

Evaluative remarks

However, from 2000 the research core funding will be separated into two parts. The first part, 80% ($80 million out of $100 million) of the total research funding from Vote Education, will be allocated through tuition fees based on student numbers in a similar way to the current system, while the remaining 20%,[31] will be allocated through a contestable pool, namely through peer review of applications. Thus, unlike the funding allocation in the UK, which is based on ex-post evaluation, the allocation is based on ex-ante evaluation. Universities have to apply for funds. Applications are assessed through peer review on the basis of the following criteria:

1. *Demonstrated quality and capacity of researchers*. The application should include the track record of the researcher in the area of research of the application. The quality and capacity of researchers will be judged on the basis of factors such as recognised publication and national and international significance of past research.
2. *Quality of the proposed research portfolio*. Quality is measured in respect of the design and purpose of the research portfolio, its feasibility, the research methods, the skills and technologies to be developed and used, and the relationship to other portfolios and programmes.
3. *Strategic focus*. The researcher should be able to demonstrate how the proposed research will increase the innovation and human resources capability in New Zealand.
4. *Cost-effectiveness*. The application is also evaluated with regard to the appropriateness of the proposed costs.

[31] These funds will be targeted at advance basic research with strategic importance. A strategically-focused research project of up to 3 years will be preferred to a short-term project.

The operation of this scheme will be reviewed in 2001. Based on the results of this review, if the contestable pool turns out to be successful in delivering high quality research in a cost-effective manner, the portion of the contestable funds will be increased gradually up to 80% over a three to five year period, thereby reducing the other part to 20%. It is expected that the Ministry will announce the exact funding mechanism by the end of October 1999 (personal communication with Paul Reynold). This will include further information on the relative weight given to each of the above criteria.

6.3 Hong Kong[32]

The higher education sector in Hong Kong comprises 3 research universities, 2 polytechnics, 2 liberal arts colleges, a teacher education institute, a performing arts conservatory and an open university. All but one of these are degree-awarding institutions (and will be referred to as universities) and are publicly funded. In addition there are 2 publicly funded technical colleges that provide 3-year higher diploma courses. All institutions enjoy full autonomy with their own Ordinance and Governance Council. Public funding for the higher education sector is provided through the University Grants Committee (UGC), a non-statutory body that for more than 30 years has acted as the principal advisor to the Government of Hong Kong on the development, funding and quality assurance of higher education.

Research in universities is funded following the dual funding system. The UGC provides the institutional core funding that covers all recurrent costs pertaining to teaching and mainstream research. In addition, the Research Grants Council (RGC), which operates under the aegis of the UGC, administers grants for academic research projects undertaken in universities. The allocation of research grants from the RGC is based on peer review involving international experts. To ensure a high degree of selectivity in the allocation of funds to high quality research, the council follows a rigorous international practice of peer review (RGC, 1998). As to the provision of the research component of the institutional core funding, it is based on the performance of university research. Since 1993, the UGC has followed the model of the UK RAE as the mechanism used in determining the research core funding allocation to universities. It has also adopted the same name for the whole evaluation process: ‘Research

[32] This section is based on French *et al*. (1998) and a publication by UGC (1999) Hong Kong on “Research Assessment Exercise 1999, Guidance Notes”.

Assessment Exercise' (RAE). Two RAEs were carried out in 1993 and 1996, and the next exercise is in 1999 (ongoing).

Evaluation performer and purposes

Initially, the UGC assigned the RGC to administer the whole process of the RAE. However in January 1993, after a preliminary study of the proposed method, it took over this responsibility although with substantial involvement of the RGC. The aim of the RAE is to provide the basis of the research funding allocation by UGC for the next three budget years. So, for the 1999 RAE it is officially stated that the purpose of the exercise is "to measure the output and quality of research of the UGC-funded institutions by cost centre as the basis for allocating some of the research portion of the institutional recurrent grant for the next triennium in a publicly accountable way"(UGC, 1999).

Evaluative remarks

Methodologically, the Hong Kong RAE is similar to that of the UK. In fact, during the first exercise in 1993, UK experts on the RAE provided an intensive consultation (French *et al.*, 1998). In general, with regard to the 1999 exercise, the method remains the same. Therefore, with regard to the method, it is sufficient to note that the unit of analysis in the evaluation is the 58 'cost centres' that are listed officially by UGC, and each department or research unit within a university is required to map their department or research unit into the cost centres in accordance with the guidelines established by the UGC. These 58 cost centres are evaluated by 12 panels (Biology, Health Sciences, Physical Sciences, Electrical & Electronic Engineering, Computer Science/Information Technology, Engineering, Built Environment, Law, Business Studies & Economics, Social Sciences & Education, Humanities and Creative Arts, Performing Arts & Design).

The ultimate aim of the 1999 RAE will be to determine the 'research index' of a cost centre, i.e. the percentage of full-time equivalent researchers in each cost centre whose research work is judged to have reached or surpassed the 'quality threshold'; in turn, this research index will be used as a basis for the allocation of funding. The research index, p, is determined by the following formula:

$$p = 100\% * A / T$$

where

T = the total number of academic staff (in FTEs) in a cost centre who meet the stated criteria, regardless of the source of funding and of whether they submit research output items for assessment; and

A = the total number of these who are judged by the Panel to have reached or surpassed the 'quality threshold'.

The formula marks the main difference between the Hong Kong RAE and the UK one. In the British RAE, the end result depends on the number of staff nominated as active researchers. As can be seen from the above formula, in the Hong Kong RAE, the end result is independent of the number of nominated academic staff because it is divided by the total number of staff (not the number nominated, as is the case with the UK RAE).

The main task for the Panel then is to determine the value of A in the above formula for each cost centre. Only those active researcher who according to the Panel reach or surpass the 'quality threshold' will be counted to add to A. The quality threshold is generally defined as *"Quality of output equates to an attainable level of excellence appropriate to the discipline in Hong Kong, and showing some evidence of international excellence"* (UGC, 1999). However, it will be up to each panel, with its subject expertise and knowledge of local circumstances, to interpret this general definition and to set up more precise benchmarks appropriate to each discipline or group of disciplines. The panels will also be expected to interpret 'international excellence' with due regard for the nature of those subjects that may have a strong local or regional focus. The panel evaluates the level of nominated active staff based on various measures of output performance (such as publications, patents, and artefacts) produced within the four-year period from 1 January 1995 to 31 December 1998.

7. Summary and Conclusions

From this survey of 18 countries, it is clear that there is increasing emphasis on research assessment. This is a consequence of constraints on public expenditure and growing demands for accountability. Funding for university research has not been exempt from these pressures, although academics are often opposed to what they see as a threat to their autonomy. Different nations have responded to these economic and political pressures to differing degrees and in different ways, some approaching the task of research assessment by relying on the traditional approach of peer review, others attempting to employ quantitative performance indicators.

The literature on research assessment is now very extensive. In this report, we have focused primarily on literature pertaining to the evaluation of university research at a national level and carried out for the purpose of informing decisions on the allocation of funds. However, in Section 2, we also looked more generally at the strengths and weaknesses of different approaches to research assessment - in particular of peer review and of bibliometric indicators. Peer review is still favoured by most researchers as the preferred approach to research assessment but it is not without its limitations. Some of these can, in principle, be partly overcome if peer review is complemented by performance indicators. However, it is difficult and often expensive to calculate the performance indicators in a sufficiently reliable and robust form to use for evaluations where the results are to be used to inform funding decisions.

In most of the 18 countries considered, some form of 'dual support' system is in operation. General infrastructure research funds are provided on an institutional basis while specific research costs are met through project grants. Three main approaches to general university funds have been identified.

1. University research funds are allocated, at least in part, on the basis of some form of research evaluation – this category can, in turn, be divided into two sub-components depending on whether the research assessment is based on (a) informed peer review (e.g. UK, Hong Kong) or (b) performance indicators (e.g. Australia)

2. University research funds are based on the education size of the university – can be sub-divided into (a) countries where university size is the sole determinant (e.g. Germany, Italy, Sweden, Norway) and those where a small proportion of

research funds is allocated on the basis of research performance (e.g. Finland, Denmark)

3. University research funds are determined through a process of negotiation between universities and the ministry responsible – can be subdivided into (a) countries in which the model is applied without any research evaluation (e.g. Austria) and (b) those where information from evaluation is taken into account (France)

   In addition, there is a fourth category of countries where:

4. Research assessment is carried out but is not linked to funding decisions – this category can again be subdivided into (a) countries where there is a dual-support system (e.g. The Netherlands) and (b) those where there is not (the USA).

   Let us consider our findings for each of these four categories of countries in turn.

7.1 General University Funds Linked to Assessment of Research Performance

Assessment based on informed peer review

The UK is probably the country with the longest experience of university research evaluation, having carried out four Research Assessment Exercises since 1986. These have benefited from incremental improvements responding to earlier criticisms, and particularly from increased transparency. The approach is based on peer review (carried out by small panels) and informed by publication and other data. The results of the assessments now influence a large proportion of the general research funding provided to universities by the Higher Education Funding Councils. In Hong Kong, the approach is very closely modelled on the UK Research Assessment Exercise although the results are used in a slightly different way in relation to the funding formula.

Assessment based on performance indicators

In determining the allocation of funds to universities, Australia makes use of the Research Quantum which is based on input data (grants and contracts) and output indicators (publication counts and higher degrees). Such an evaluation system is cheaper to operate than the UK Research Assessment Exercise but faces some potential problems arising, for example, as a result of the variation in research costs across and

even within fields. In the Slovak Republic, university departments (like academy institutes) are now assessed using a number of indicators, and the results influence the research funding they receive. In Poland, the assessment of universities was previously based on peer review but in 1998 a new formula-based approach was adopted, the formula being based on such indicators as numbers of publications, scientific degrees and research income.

7.2 General University Funds Based on University Size

No account taken of research performance

In several of the countries analysed here, research funding is allocated to universities as part of the general institutional funding which is based on university size (numbers of students and staff). In such countries (as in those listed under the first category above), universities also receive research funds for specific projects from research councils, which employ peer review to assess the proposals that they receive.

In Germany, there has been no comprehensive assessment of university research, although the state of Lower Saxony recently started to assess certain fields in its universities, and other Länder governments are beginning to allocate a proportion of research funding on a performance-related basis. In addition, the Freie University in Berlin has implemented its own evaluation system, the results of which are used in determining the internal distribution of funds.

In Norway and Sweden, universities receive a block grant based on their size with no attempt made to assess research performance. The same has been true for Italy but the National University Evaluation Council has recently proposed that funding be linked, among other things, to research results.

Small proportion of research funds allocated on basis of research performance

In Finland, universities receive most of their general research support as part of their institutional funding which is determined by university size. However, a small proportion of this funding is based on their performance in teaching and research (e.g. numbers of PhD degrees). Likewise, in Denmark, since 1995, a small element of university funding has been linked to levels of external research income and numbers of PhDs produced. Universities in New Zealand have previously received research funding through the general institutional support (based on student numbers) but in 2000 there will be a switch to contested funding allocated on the basis of *ex ante* evaluation

(carried out by peer review). The proportion of research funds allocated to universities in this way may increase over time if the approach proves successful.

7.3 General University Funds Based on Negotiation with Ministry

No research evaluation

In Austria as in France (see below), university budgets are determined primarily through a process of negotiation with the relevant ministry. In the case of Austria, there is no strict relation between university size and funding in arriving at decisions on funding, although this is taken into account in the negotiation process. No research assessment is involved. The academic community is strongly opposed to such assessments and there has been no strong push in this direction from ministers. Evaluations of two fields have been carried out but the results had no formal consequences.

Some account taken of university evaluations

In France, the negotiation of university funding levels involves not only the Ministry of Education, Research and Technology and the universities but also the main research council, CNRS. Unlike Austria, however, in France these negotiations take into account the evaluations carried out of individual universities by the National Evaluation Committee. These evaluations focus on teaching and management as well as research, and they are based on peer review along with various indicators.

7.4 Research Assessments Conducted but not Linked to Funding

Countries with a dual-support system

There is extensive evaluation of university research in The Netherlands but it is not used for decisions on the distribution of funding among universities. Instead, the aim of such evaluations is to help those responsible for the research management of universities, for example in developing research strategies and in decision-making. The evaluation approach is based on informed peer review carried out by panels (as in the UK) but with very heavy reliance on foreign peers and some site visits. Another difference compared with the UK is that different fields are evaluated at different times (rather than all being done together). In addition, the evaluations focus on not one but four aspects of research performance – scientific quality, scientific productivity, scientific relevance and long-term viability.

In Canada, the first extensive research assessment of universities is being carried out in 1999. As in The Netherlands, the results are not being used to decide on the allocation of funds to universities. Instead, they will go to individual universities to be used for their own purposes.

Countries without a dual-support system

In the United States, universities receive almost all their research income in the form of specific research grants and contracts. The institutional funding that public universities receive from state governments is primarily for teaching purposes. Nevertheless, there have been periodic evaluations of all research doctorate programmes in universities. These assess the research standing of departments as well as the quality of their doctoral training. The approach is based on peer review but in this case it takes the form of an extensive questionnaire survey rather than a small panel (as in the UK and The Netherlands, for example). Use is also made of various performance indicators. The results are employed by doctoral students (and perhaps by academic staff) in choosing which department to join, and by universities in relation to internal management and strategy.

Let us conclude by stressing two points. The first is that, although only a small number of countries are currently using performance-based approaches to determine the allocation of university research funding, most of those considered in this report will soon implement some form of performance-based allocation or are considering doing so.

Secondly, this study did not aim to provide a better understanding of the implications of the Research Assessment Exercise (RAE) for higher education institutions in England, nor to make an assessment of possible policy alternatives facing the Higher Education Funding Council for England. As such, this report provides a description of current practice in research assessment at the national level for the purpose of informing the funding allocation process in a large number of countries. It does not provide an analysis of the efficiency and efficacy of the different funding approaches.

Recommendations/Action Items

Finally, what action might the Higher Education Funding Council for England consider taking in the future? We would suggest that consideration be given to the following:

- HEFCE should continue to monitor current practice in university assessment and funding allocation procedures in other countries.
- HEFCE should assess the intended and unintended consequences of the current funding approach used in England, and compare this to one or more of the approaches adopted in other countries.

Bibliography

Adams, J. D. and Z. Griliches. 1996. *Research Productivity in a System of Universities*. Cambridge: National Bureau of Economic Research.

Arora, A., P. David, and A. Gambardella. 1998. Reputation and Competence in Publicly Funded Science: Estimating the Effects on Research Group Productivity. *Annales d'Economie et de Statistique*, Vol. 49/50: pp. 163-198.

Atkinson, H. and W. F. Massy. 1996. Quantitative Funding Models. In *Resource Allocation in Higher Education*, ed. William F Massy. Michigan: The University of Michigan Press.

Baimbridge, M. 1998. Institutional Research Performance 1992-96: A Tale of Two Sectors. *Journal of Further and Higher Education*, Vol. 22(1): pp. 69-78.

Ball, D. F. 1997. Quality Measurement as a Basis for Resource Allocation: Research Assessment Exercise in United Kingdom Universities. *R&D Management*, Vol. 27(3): pp. 281 - 289.

Beasley, J. 1990. Comparing University Departments. *OMEGA Int. J. org Mgmt Sci.*, Vol. 18(2): pp. 171-183.

Biggeri, L. and L. Scarpitti. 1998. Evaluation in the Italian University System. In *International Conference on Evaluation: Profession, Business or Politics?* Rome, October 29-31, 1998.

BMBF. 1998. *Facts and Figures 1998*. Bonn: Bundesministerium für Bildung und Forschung (BMBF).

Boffo, S. 1996. *EVALUE, State of Play Report: Italy*. Contract SOE2 - CT 95 - 2004. European Commission, DG XII, Programme TSER (Targeted Socio-Economics Research).

Bourke, P. 1997. *Evaluating University Research: The British Research Assessment Exercise and Australian Practice*. Commissioned Report No. 56. Canberra: National Board of Employment, Education and Training.

Bourke, P. and B. Martin. 1992. *Evaluating University Research Performance - What Approach? What Unit of Analysis*. ANU & SPRU.

Brennan, M. 1994. Excellence and Relevance - Two Sides of the Same Coin. *Higher Education*, Vol. 28: pp. 129-135.

Brofoss, K. E. 1998. The Research Council of Norway's Use of Research Evaluation: An Assessment of Research Evaluation as a Strategic Tool. *Research Evaluation*, Vol. 7(3): pp. 134-140.

Caraça, J., P. Conceição, and M. V. Heitor. 1998. A Contribution towards a Methodology for University Public Funding. *Higher Education Policy*, Vol. 11: pp. 37-57.

Carpenter, M. P., F. Gibbs, M. Harris, J. Irvine, B. R. Martin, and F. Narvin. 1988. Bibliometric Profiles for British Academic Institutions: An Experiment to Develop Research Output Indicators. *Scientometric,* Vol. 14(3-4): pp. 213 - 233.

Cave, M., S. Hanney, M. Henkel, and M. Kogan. 1997. *The Use of Performance Indicators in Higher Education.* Higher Education Policy Series. London: Jessica Kingsley Publisher.

Chubin, D. E. 1987. Research Evaluation and the Generation of Big Science Policy. *Knowledge: Creation, Diffusion, Utilization,* Vol. 9(2): pp. 254 - 277.

Cole, J. and S. Cole. 1972. The Ortega Hypothesis. *Science,* Vol. 178: pp. 368-375.

Cooper, C. and D. Otley. 1998. The 1996 Research Assessment Exercise for Business and Management. *British Journal of Management,* Vol. 9: pp. 73-89.

Cozzens, S. 1995. U.S. Research Assessment: Recent Development. *Scientometric,* Vol. 34(3): pp. 351-362.

Cozzens, S. E. 1989. What do Citations Count? The Rhetoric-First Model. *Scientomeric,* Vol. 15: pp. 437 - 447.

Daniel, H. D. and R. Fisch. 1990. Research Performance Evaluation in the German University Sector. *Scientometrics,* Vol. 19(5-6): pp. 349-361.

Department for Education and Science Policy. 1997. *Management by Results in Higher Education*. Department for Education and Science Policy/Universities Branch, Ministry of Education, Science and Culture, Finland.

DETYA: Department of Education Training and Youth Affairs. 1998. *The Composite Index Allocation of the Research Quantum to Australian Universities*. Canberra, Australia: Department of Education, Training and Youth Affairs.

Ehrenberg, R. G. and P. J. Hurst. 1996. The 1995 NRC Ratings of Doctoral Programs A Hedonic Model. *Change,* Vol. 28(3): pp. 46-50.

El-Khawas, E. and W. F. Massy. 1996. Britain's "Performance-Based" System. In *Resource Allocation in Higher Education,* ed. William F. Massy. Michigan: University of Michigan Publisher.

Fairweather, J. 1988. Reputational Quality of Academic Programs: The Institutional Halo. *Research in Higher Education,* Vol. 28(4): pp. 345-355.

Felderer, B. and D. F. J. Campbell. 1998. *Evaluation der Akademischen Forschung*. Vienna: Institute for Advanced Studies.

Frankel, M. S. and J. Cave. 1997. Introduction. In *Evaluating Science and Scientist: An East-West Dialogue on Research Evaluation in Post-Communist Europe*, ed. M.S. Frankel and J. Care. Budapest: Central European University Press.

French, N. J., W. F. Massy, P. K. Ko, H. F. H. Siu, and K. Young. 1998. Research Assessment in Hong Kong. *Journal of International Education,* Vol. 10(1).

Gellert, C., ed. 1993. *Higher Education in Europe*. London and Philadelphia: Jessica Kingsley Publisher.

Geuna, A. 1999a. *The Economic of Knowledge Production: Funding and the Structure of University Research*. Cheltenhem: Edward Elgar.

Geuna, A. 1999b. *The Changing Rationale for European University Research Funding: Are there Negative Unintended Consequences?* Working Paper. Brighton: SPRU.

Gillett, R. 1989. Pitfalls in Assessing Research Performance by Grant Income. *Scientometrics,* Vol. 22(2): pp. 253-263.

Goldberger, M. L., B. A. Maher, and P. E. Flattau. 1995. *Research-Doctorate Programs in the United States*. Washington, D.C.: National Research Council.

Griffith, J. 1995. *Research Assessment: As Strange a Maze as E'er Men Trod*. Report No. 4. London: Council for Academic Freedom & Academic Standards.

Guy-Ohlson, D. 1997. International Evaluations of the Swedish Natural Science Research Council (NFR). In *The Evaluation of Scientific Research: Selected Experiences*, ed. OECD. Paris: OECD.

Hangos, K. 1997. The Limits of Peer Review: The Case of Hungary. In *Evaluating Science and Scientist: An East-West Dialogue on Research Evaluation in Post-Communist Europe*, ed. M.S. Frankel and J. Care. Budapest: Central European University Press.

Hanson, R. 1994. Allocation and Evaluation: The Approach at the Social Sciences and Humanities Research Council of Canada. *Higher Education,* Vol. 28: pp. 109-117.

Hare, P. and G. Wyatt. 1988. Modelling the Determination of Research Output in British Universities. *Research Policy,* Vol. 17: pp. 315 - 328.

Harris, G. and G. Kaine. 1994. The Determinants of Research Performance: A Study of Australian University Economist. *Higher Education,* Vol. 27: pp. 191 - 201.

HEFCE. 1997. *A Guide to Funding Higher Education in England*. Accessed 23 September 1997. http://www.niss.ac.uk/education/hefce/funding/index.htm.

Helander, E. 1995. Evaluation Activities in the Nordic Countries. *Scientometrics,* Vol. 34(3): pp. 391-400.

Hemlin, S. 1998. Utility Evaluation of Academic Research: Six Basic Proposition. *Research Evaluation,* Vol. 7(3): pp. 159-165.

Hills, P. V. and A. J. Dale. 1995. Research and Technology Evaluation in the United Kingdom. *Research Evaluation,* Vol. 5(1): pp. 35-44.

Industry Commission. 1995. *Research and Development*. Report No. 44, Volume 1. Canberra: Assistant Treasurer.

Irvine, J., B. R. Martin, and P. A. Isard. 1990. *Investing in the Future, An International Comparison of Government Funding of Academic and Related Research.* Hants: Edward Elgar.

Jablecka, J. 1995. Changes in the Management and Finance of the Research System in Poland: A Survey of the Opinions of Grant Applicants. *Social Studies of Sciences,* Vol. 25: pp. 727-753.

Jenkins, A. 1995. The Impact of the Research Assessment Exercise on Teaching in Selected Geography Departments in England and Wales. *Geography,* Vol. 80 part 4(349): pp. 367-374.

Johnes, G. 1988. Determinants of Research Output in Economics Departments in British Universities. *Research Policy,* Vol. 17: pp. 171 - 178.

Johnes, G. 1990. Measures of Research Output: University Departments of Economics in the UK, 1984-8. *Economic Journal,* Vol. 100: pp. 556-560.

Johnes, G. 1992. Performance Indicators in Higher Education: A Survey of Recent Work. *Oxford Review of Economic Policy,* Vol. 8(2): pp. 19 - 34.

Johnes, G. and J. Johnes. 1993. Measuring the Research Performance of UK Economics Departments: An Application of Data Envelopment Analysis. *Oxford Economic Papers,* Vol. 45: pp. 332 - 347.

Johnes, J. and J. Taylor. 1992. The 1989 Research Selectivity Exercise: A Statistical Analysis of Differences in Research Rating between Universities at the Cost Centre Level. *Higher Education Quarterly,* Vol. 46(1): pp. 67-87.

Johnes, J., J. Taylor, and B. Francis. 1993. The Research Performance of UK Universities: A Statistical Analysis of the Results of the 1989 Research Selectivity Exercise. *Journal of the Royal Statistical Society - A,* Vol. 156(2): pp. 271-286.

Kaa, D. J. V. d. 1994. Picking the Winners by Consensus: Grant-giving Practice in The Netherlands. *Higher Education,* Vol. 28: pp. 59 - 83.

Kaukonen, E. 1997. Evaluation of Scientific Research in Finland. In *The Evaluation of Scientific Research: Selected Experiences,* ed. OECD. Paris: OECD.

Kennedy, D. 1986. Basic Research in the Universities: How Much Utility? In *The Positive Sum Strategy: Harnessing Technology for Economic Growth,* ed. Landau and R. Rosenberg. Washington D.C.: National Academic Press.

Koelman, J. B. J. 1998. The Funding of Universities in The Netherlands: Developments and Trends. *Higher Education,* Vol. 35: pp. 127-141.

Kogan, M. 1989. The Evaluation of Higher Education: An Introductory Note. In *Evaluating Higher Education,* ed. Maurice Kogan. London: Jessica Kingsley Publisher.

Kogan, M. 1998. The Treatment of Research. *Higher Education Quarterly*, Vol. 52(1): pp. 48-63.

Kuhlman, S. 1995. Pattern of Science and Technology Policy Evaluation in Germany. *Research Evaluation*, Vol. 5(1): pp. 23-33.

Kushner, S. 1996. The Research Assessment Exercise versus Development in Higher Education: A Response to Richard Pring. *British Journal of Educational Studies*, Vol. 44(1): pp. 5-8.

Kyvik, S. 1997. Funding University Research in the Nordic Countries. *Science and Public Policy*, Vol. 24(4): pp. 233-244.

Laredo, P. 1998. *Changing Structure, Organisation and Nature of PSR Systems*. Research funded in part by the European Commission TSER SOE1-CT96-1036. Centre de Sociologie de l'innovation Ecole Nationale Superieure des Mines de Paris.

Lopez-Martinez, R. E. and A. Rocha-Lackiz. 1998. University R&D Performance Evaluation: The Case of the National University of Mexico. *Research Evaluation*, Vol. 7(3): pp. 167-177.

Luukkonen, T. 1995. The Impacts of Research Field Evaluations on Research Practice. *Research Policy*, Vol. 24: pp. 349-365.

Maher, B. A. 1996 Nov-Dec. The NRC's Report on Research-Doctorate Programs Its Uses and Misuse. *Change*, Vol. 28(6): pp. 54-59.

Martin, B. and J. Irvine. 1983. Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy. *Research Policy*, Vol. 12: pp. 61-90.

Martin, B. R. 1996. *The Use of Multiple Indicators in The Assessment of Basic Research*. Brighton: SPRU.

Martin, B. R. 1997. Factors Affecting the Acceptance of Evaluation Results. In *Evaluating Science and Scientist: An East-West Dialogue on Research Evaluation in Post-Communist Europe*, ed. M.S. Frankel and J. Care. Budapest: Central European University Press.

Martin, B. R. and J. E. F. Skea. 1992. *Academic Research Performance Indicators: An Assessment of the Possibilities*. Brighton: SPRU.

Massy, W. F. 1996. *Resource Allocation in Higher Education*. Michigan: The University of Michigan Press.

McNay, I. 1999. The Paradoxes of Research Assessment Funding. In *Changing relationship between higher education and the state*, ed. Mary Henkel and Brenda Little:191 - 203. London: Jessica Kingsley Publishers.

Meer, P. v. d. 1999. *Funding and Allocation of Resources in Higher Education: The Ducth Case*. Paper prepared for the 3rd ALFA_BRACARA international

conference on 'Funding and Allocation of Resources in Higher Education' 8-10 February 1999 at Universidaded Nacional Autonoman do Mexico.

MES. 1979. *Beleidsnota Universitair Onderwijs*. The Hague: Minister for Education and Science, Minister for Science Policy and Minister for Agriculture and Fisheries.

Ministry of Education. 1998. *Tertiary Education in New Zealand, Policy Direction for the 21st Century*. Wellington: Ministry of Education of New Zealand.

Moed, H., M. Luwel, J. Houben, E. Spruyt, and H. V. D. Berghe. 1998. The Effects of Changes in the Funding Structure of the Flemish Universities on Their Research Capacity, Productivity and Impact during the 1980's and Early 1990's. *Scientometric,* Vol. 43(2): pp. 231-255.

Moed, H. F., W. J. M. Burger, J. G. Frankfort, and A. F. J. v. Raan. 1985. The Use of Bibliometric Data for the Measurement of University Research Performance. *Research Policy,* Vol. 14: pp. 131 - 149.

Mora, J.-G., J. Palafox, and F. Perez. 1995. The Financing of Spanish Public Universities. *Higher Education,* Vol. 30: pp. 389 - 406.

Murphy, P. S. 1995. Benchmarking Academic Research Output in Australia. *Assessment & Evaluation in Higher Education,* Vol. 20(1): pp. 45 - 57.

Nederhof, A. J. and E. C. M. Noyons. 1992. Assessment of the International Standing of University Department's Research: A Comparison of Bibliometric Methods. *Scientometrics,* Vol. 24: pp. 393 - 404.

Nederhof, A. J. and A. F. J. v. Raan. 1993. A Bibliometric Analysis of Six Economies Research Groups: A Comparison With Peer Review. *Research Policy,* Vol. 22: pp. 353 - 368.

OECD. 1987. *Universities Under Scrutiny*. Paris: OECD.

OECD. 1990. *Financing Higher Education*. Paris: OECD.

OECD. 1997. *The Evaluation of Scientific Research: Selected Experiences*. Paris: OECD.

OECD. 1998. *University Research in Transition*. Paris: OECD.

Phillimore, A. J. 1989. University Research Performance Indicators in Practice: The University Grants Committee's Evaluation of British Universities, 1985-86. *Research Policy,* Vol. 18: pp. 255 - 271.

Print, M. and J. Hattie. 1997. Measuring Quality in Universities: An Approach to Weighting Research Productivity. *Higher Education,* Vol. 33: pp. 453-469.

Ran, R. T. H. v., ed. 1988. *Handbook of Quantitative Studies of Science and Technology*. Amsterdam.

RGC. 1998. *Refutation of Article in "Next" Magazine of 9 October 1998*. Accessed 20/7/99 1999. http://ww.ug.edu.hk/RGC/documents/E_XNEXT.html.

Rip, A. and B. J. R. v. d. Meulen. 1995. The Patchwork of the Dutch Evaluation System. *Research Evaluation,* Vol. 5(1): pp. 45-53.

Royal Society of Chemistry. 1995. *Chemistry in the UK - Will it Survive? Conclusions of the Royal Society of Chemistry Workshop*. London: Royal Society of Chemistry.

Schweighofer, M.-G. 1997. Research Evaluation at the Centre National de la Recherche Scientifique (CNRS) in France. In *The Evaluation of Scientific Research: Selected Experiences,* ed. OECD. Paris: OECD.

Senker, J. 1999. *University Research Financing, A Report prepared for the OECD*. Brighton: Science Policy Research Unit (SPRU).

Sizer, J., A. Spee, and R. Bormans. 1992. The Role of Performance Indicators in Higher Education. *Higher Education,* Vol. 24: pp. 133-155.

Smith, T. 1987. The UGC's Research Ranking Exercise. *Higher Education Quarterly,* Vol. 41: pp. 303-316.

Steen, J. v. and M. Eijffinger. 1998. Evaluation Practices of Scientific Research in The Netherlands. *Research Evaluation,* Vol. 7(2): pp. 113-122.

Steiner, M. and D. Sturn. 1995. Elements of Evaluation of Science and Technology Policy in Austria. *Research Evaluation,* Vol. 5(1): pp. 98-108.

Suter, L. E. 1997. United States: The Experience of the NSF's Education and Human Resources Directorate. In *The Evaluation of Scientific Research: Selected Experiences,* ed. OECD. Paris: OECD.

TFR. 1997. *Peers on Peers: Allocation Policy and Review Procedures at the Swedish Research Council for Engineering Sciences*. TFR Evaluation Program, Volume 4. Stockholm: Swedish Research Council for Engineering Sciences (TFR).

THES: The Times Higher Education Supplement, 20 December 1996. Supplement Research Assessment Exercise 1996.

THES: The Times Higher Education Supplement, 28 February 1997, pp. 16-17. Assessment Reassessed.

THES: The Times Higher Education Supplement, 14 November 1997, p. 12. RAE Goes Forth and Multiplies.

Tino, J. 1997. Institutionalizing Evaluation of Research and Education in the Slovak Republic. In *Evaluating Science and Scientist: An East-West Dialogue on Research Evaluation in Post-Communist Europe,* ed. M.S. Frankel and J. Care. Budapest: Central European University Press.

Tomlin, R. 1998. Research League Tables: Is There a Better Way? *Higher Education Quarterly,* Vol. 52(2): pp. 204-220.

UFC. 1989. *Report on the 1989 Research Assessment Exercise*. London: University Funding Council.

UGC. 1999. *Research Assessment Exercise 1999, Guidance Notes*. University Grant Committee Hong Kong.

VSNU. 1996. *Quality Assessment of Research: Chemistry, Past Performance and Future Perspectives*. Utrecht: VSNU, Association of Netherlands Universities.

Vught, F. A. v. and D. F. Westerheijden. 1994. Towards a General Model of Quality Assessment in Higher Education. *Higher Education,* Vol. 28: pp. 355-371.

Webster, D. S. 1996 May-June. Rating PhD Programs: What the NRC Reports Says...and Doesn't Say. *Change,* Vol. 28(3): pp. 34-44.

Westerheijden, D. F. 1997. A Solid Base for Decisions, Use of the VSNU Research Evaluations in Dutch Universities. *Higher Education,* Vol. 33: pp. 397-413.

Whittington, G. 1997. The 1996 Research Assessment Exercise. *British Accounting Review,* Vol. 29: pp. 181-197.

Williams, B. 1993. Research Management for Selectivity and Concentration with What Effect? *Higher Education Quarterly,* Vol. 47(1): pp. 4-16.

Williams, D. and A. D. Rank. 1998. Measuring the Economic Benefits of Research and Development: The Current State of the Art. *Research Evaluation,* Vol. 7(1): pp. 17 - 30.

Wolfe, D. A. 1998. *The Role of the Provinces in PSE Research Policy*. Ontario, Canada: Discussion Paper prepared for the Council of Ministers of Education.

Wood, F. Q. 1995. Issues and Problems in the Public Funding of University Basic Research. PhD Thesis, University of New England.

Wood, F. Q. 1997. *The Peer Review Process*. Commissioned Report No. 54. Canberra: National Board of Employment, Education and Training, Australian Research Council.

Wood, F. Q., V. L. Meek, and G. Harman. 1992. The Research Grant Application Process. Learning from Failure? *Higher Education,* Vol. 24: pp. 1-23.

Zilahy, P. and I. Láng. 1997. The Evaluation of Research Institutes in Hungary. In *Evaluating Science and Scientist: An East-West Dialogue on Research Evaluation in Post-Communist Europe,* ed. M.S. Frankel and J. Care. Budapest: Central European University Press.

Appendix 1: Sources of Personal Communication

1. Dr. Satis G. Arnold
   Director, Research Policy
   Higher Education Division
   Department of Employment, Education, Training and Youth Affairs
   GPO Box 9880, Canberra ACT 2601, Australia

2. Dr. Rémi Barré
   Directeur, Office of Science and Technology (OST)
   93 rue de Vaugirard 75006, Paris

3. Dr. Julita Jablecka
   Centre for Science Policy and Higher Education Studies,
   Warsaw University, Nowy Swiat 69, Poland

4. Dr. Anne Klemperer
   Center for Higher Education Policy Studies (CHEPS)
   University of Twente
   PO Box 217, 7500 AE Enschede, The Netherlands

5. Dr. Paul Reynold
   Chief Policy Adviser
   Ministry of Research, Science and Technology
   P O Box 5336, Wellington, New Zealand

6. Ing. Štefan Zajac, CSc.
   Slovenská akadémia vied
   Prognostický ústav, Šankcová 56, the Slovak Republic.