Kommentierte Bibliographie zur Forschungs- und Hochschulevaluation in ausgewählten Ländern

Hans-Dieter Daniel

Center for Science and Technology Studies
Zentrum für Wissenschafts- und Technologiestudien
Centre d'études de la science et de la technologie
Centro di studi sulla scienza e la tecnologia

CEST 2001/2a

Inhaltsverzeichnis

Kommentierte Bibliograhie zur Forschungs- und Hochschulevaluation in ausgewählten Ländern

von Hans-Dieter Daniel

Kommentierte Bibliographie zur Forschungs- und Hochschulevaluation in ausgewählten Ländern

1. Kanada

Forschungsevaluation

Agora Management Associates (1999). A Profile of Audit, Evaluation and Review Units in the Federal Government, 1999.
http://www.ncr.dfo.ca/COMMUNIC/cread/english/reports/aer/aer.htm

The purpose of the study is to give Review Network members a snapshot of how their function is organized and resourced at present; of significant characteristics and issues regarding the mission, policies and practices in this area; and of the effects of the environment on their work (The Network is made up of many of the heads of audit, evaluation and review in the Federal Government of Canada).

Anderson, Frances & Dalpé, Robert (1996). S&T indicators for strategic planning and assessment of public research institutions. Knowledge and Policy: The international journal of knowledge transfer and utilization, 9(1), 49-69.

Anderson, Frances & Dalpé, Robert (1991). The evaluation of public applied research laboratories. Canadian Journal of Program Evaluation, 6(2), 107-125.

Anderson, Frances & Gault, Fred (1999). Developing indicators of the internationalization of R&D: the case of Canada. Research Evaluation, 8(1), 15-22.

The paper analyses foreign sources of funding for Canadian R&D performers and shows that 20% of the funding for Canadian R&D in the business sector comes from foreign sources which, when compared to other countries, is among the highest percentages. Nevertheless, very few firms receive foreign R&D funding and even less on a regular basis. Firms tend to have transactions with foreign affiliated firms or foreign non-affiliated firms but rarely engage in both types of transaction. The rise in foreign funding of R&D in domestic firms suggests that there may be recognition to the favourable tax environment for Canadian firms.

ARA Consulting Group (1997). Evaluation of the networks of Centres of Excellence Programme: Final Report (prepared for the NCE Programme Evaluation Committee). Vancouver, British Columbia: ARA Consulting Group, Inc.

Auditor General (1993). Program Evaluation in the Federal Government. The Case for Program Evaluation. Treasury Board of Canada.
http://www.oag-bvg.gc.ca/domino/reports.nsf/html/ch9308e.html

Program evaluation is a disciplined assessment of government programs and activities. It is based on independent, systematic measurement and analysis, carried out to meet expectations set in policy and standards, and publicly reported. The first government policy requiring program evaluation was issued in 1977. It followed in a tradition of administrative reforms that attempted to link the systematic collection and analysis of information to improved public sector management. Six years after the first evaluation policy, our 1983 audit found program evaluation in place in most departments in the federal government. Over the next ten years, the development of the function took place largely in these departments rather than in the central agencies.
There were high expectations for program evaluation. It was expected to support key functions of government - allocation of resources, efficient and effective operations, and accountability reporting. In addition, program evaluation was expected to serve many different information needs,

including those of managers of government programs, managers in central agencies, Parliament and the public.
Program evaluation is intended to have a strategic focus that periodically assesses the performance of policies and programs, providing results that can be used reliably by decision makers. Program evaluation involves applying systematic research methods drawn from many different disciplines to assess performance, particularly effectiveness. The evaluation results are presented in a report. By answering questions about what has been accomplished, program evaluation can provide an important measure of the value obtained from government programs and expenditures.
The government established program evaluation to identify whether its initiatives have been successful, whether they should be continued, or whether there are more cost-effective alternatives. The federal government is responsible for large expenditures on programs - with a budget of $161 billion for 1993-94. The expenditures involve over 1500 components, each with its own objectives. In addition, the government has in place a large number of tax expenditures and selective tax measures that also have specific purposes. Program evaluation can be the source of objective information on the complex operations of government, and can provide a measure of the value being obtained.

Auditor General (1993). The Program Evaluation System Making it Work. Treasury Board of Canada.
http://www.oag-bvg.gc.ca/domino/reports.nsf/html/ch9310e.html

Barbarie, Alain (1993). Evaluating Federal R&D in Canada. In Barry Bozeman & Julia Melkers (Eds.), Evaluating R&D Impacts: Methods and Practice (155-162). Norwell, MA: Kluwer Academic Publishers.

Dalpé, Robert (1997). International activities of public laboratories in Canada. Technology in Society, 19(2), 127-143.

Dalpé, Robert & Bélot, Jean-Marc (1999). Measuring technology through bibliometrics. Research Evaluation, 8(1), 23-31.

The possibility is examined of diffusing and adapting bibliometric methods to technical literature for the measurement of technological output. The first part of the paper deals with existing methods in the treatment of such literature for the generation of technology indicators, which consists mainly in the extraction of new products information from technical and trade journals. The second part describes the construction of a new-products database for the mechanical industry.

Dalpé, Robert & Gauthier, Élaine (1993). Evaluation of the industrial relevance of public research institutions. Research Evaluation, 3(1), 43-54.

Data is derived from the four most used data sources in public research evaluation: publications, patents, university-industry contracts, and questionnaires. The study measures the industrial relevance of Canadian academic research in metals and alloys. No databases offer the possibility of covering the whole range of possible interactions between public institutions and industry; multi-indicators are therefore needed.

Dalpé, Robert & Longpré, B. (1995). The state of Canadian research in physics and electrical engineering. Report to the NRCD, Montréal, Centre interuniversitaire de recherche sur la science et la technologie (CIRST).

D'Aoust, Raymond & Lemaire, Donald (1994). Untangling the Gordian knot: bridging instrumental rationality and stakeholder politics in the evaluation of public policies. Research Evaluation, 4(1), 37-44.

The web of public policy design and implementation in Canada is likened to the Gordian knot, which ties together the specific use of instruments and stakeholder politics. Two analytical frameworks are presented which offer great potential for untangling the knot - the instruments approach and Lundquist's concept of stakeholder politics. There is an urgent need for a

'transdisciplinary' approach to evaluation. Policy evaluators will have to use concepts and develop applications drawn from different disciplines.

Dawson, P., Dalpé, R., Longpré, B. & Caron, C. (1996). A bibliometric view of the state of Canadian research in semiconductors and photonics. La Physique au Canada, July/August, 151-158.

Finn, P. J. (1988). Evaluation of the Crop Production Development Research Program. Canadian Farm Economics, 21, 19-27.

Freedman, Ron (1993). Necessary Condition Analysis: a new tool for program evaluation. Research Evaluation, 3(2), 127-131.

There is an increasing call to evaluate programs early on in their life before any results have been obtained. A new type of assessment has been developed for this situation called Necessary Condition Analysis. Rather than evaluating results, this checks whether the necessary conditions are in place for the results to be achieved at the end of the program. Evaluators cannot be sure of success if these conditions are met but they can be sure that the program will not succeed if they are not present.

Gault, Fred D. (1999). A five-year strategic plan for the development of an information system for science and Technology. Ottawa, Ontario: Statistics Canada, Science and Technology Redesign Project.
http://www.statcan.ca:80/english/IPS/Data/88-523-XIE.htm

This publication outlines a five-year strategic plan for the development of an information system for science and technology.

Gault, Fred D. (1999). Science and technology activities and impacts: a framework for a statistical information. Ottawa, Ontario: Statistics Canada, Science and Technology Redesign Project.
http://www.statcan.ca:80/english/IPS/Data/88-522-XIE.htm

The framework described here is intended as a basic operational instrument for systematic development of statistical information respecting the evolution of science and technology and its interactions with the society, the economy and the political system of which it is a part.

Gault, Fred D. (1998). Research and development in a service economy. Research Evaluation, 7(2), 79-91.

Canada has a service economy and R&D in Canada is mainly a service sector activity. The paper examines the sectoral distribution of expenditure on R&D performance, with emphasis on the business sector in Canada and with international comparisons. Human resources are a key component in the performance of R&D, and comparisons are made, over time, of the number of research workers in service and non-service industries, of the ratio of professional to technical and other personnel, and of the changes in educational levels of R&D personnel. Using Canadian experiences as a guide, some conclusions are drawn about the measurement challenges in producing indicators of the transition to a service economy.

Garfield, Eugene (1993). What citations tell us about Canadian research. Canadian Journal of Information and Library Science, 18(4), 14-35.

The paper illustrates the variety of perspectives on Canadian science that are possible with citation data. The examples range from the general to the specific, including rankings of nations, institutions, fields, papers, and authors.

Gauthier, Élaine (1998). Bibliometric analysis of scientific and technological research: A user's guide to the methodology. Ottawa: Statistics Canada, Science and Technology Redesign Project, ST-98-08.
http://www.statcan.ca/english/research/88F0006XIB/98010.pdf

The paper provides an overview on current usage of bibliometric methods and techniques, including an extensive bibliography. It also provides technical specifications on the database of Canadian authors that has been developed, with „Statistics Canada“ support, by the „Observatoire des Sciences et des Technologies“. The paper is a companion document to two other working papers. The first, „Knowledge Flows in Canada as Measured by Bibliometrics“, uses the database to develop statistical indicators of knowledge flow in the natural sciences and engineering. The second, „The Use of Bibliometric Data to Measure Scientific Production in the Arts, Humanities and Social Sciences: A Methodological Note“, examines the issues involved in the use of bibliometrics for the social sciences, arts and humanities. Both of these working papers are authored by B. Godin, Y. Gingras and L. Davignon of the „Observatoire des Sciences et des Technologies“.

Gingras, Y. (1996). Bibliometric analysis of funded research: A feasibility study. Report to the Program Evaluation Committee of NSERC, Montréal, Centre interuniversitaire de recherche sur la science et la technologie (CIRST).

Girard, Lucie (1993). Evaluation at Fonds FCAR: discoveries and questions. Research Evaluation, 3(3), 167-172.

Two years ago Fonds FCAR (Fonds pour la Formation de Chercheurs et l'Aide à la Recherche) instated a two-prong evaluation system: a horizontal evaluation of its program conducted by staff, and an individual program evaluation undertaken by a consultant firm. The strategic choices of the agency are taken into account grouped according to five trends: maintenance of a broad equilibrium (by sector and forms of research); organisation of the research community to reach critical masses; support of new scientists in universities; the enhancement of research training; and the support of excellence. The evaluation assesses the relevance of the chosen objectives, the efficacy of the programs and their impact on the research system, and the efficiency of the chosen means.

Godin, B. (1997). Research and the practice of publication in industry. Research Policy, 25, 587-606.

Godin, B. (1997). Profil bibliométrique de la recherche financée en sciences naturelles, génie et sciences biomédicales. Rapport de recherche présenté au Fonds FCAR, Montréal, INRS/ Centre interuniversitaire de recherche sur la science et la technologie (CIRST).

Godin, B., Gingras, Y. & Davignon, L. (1998). Knowledge Flows in Canada as Measured by Bibliometrics. Ottawa: Statistics Canada, Science and Technology Redesign Project, ST 98-10.
http://www.statcan.ca/english/research/88F0006XIB/98010.pdf

Hansen, R. (1994). Allocation and evaluation: The approach at the Social Sciences and Humanities Research Council of Canada. Higher Education, 28, 109-117.

Helbing, Caren C., Verhoef, Marja J. & Wellington, Cheryl L. (1998). Finding identity and voice: a national survey of Canadian postdoctoral fellows. Research Evaluation, 7(1), 53-60.

An interdisciplinary, national study of postdoctoral fellows (PDFs) in Canada and of Canadian PDFs abroad was undertaken. The questionnaire-based study determined basic demographics, evaluated job stress/satisfaction and work environment, and assessed outlook for future careers using a variety of parameters. Although several aspects were rated satisfactory, the PDFs surveyed identified key areas of conern in their training.

Lemaire, Donald (1995). Reviewing Regulatory Programs. Guide - Review 95-03-01. Ottawa: Office of the Comptroller General, Treasury Board of Canada.
http://www.tbs-sct.gc.ca/rin/ia_main/auditguidance/REG_GDE4.e.html

The purpose of this document is to outline the value of regulatory reviews such that there use may aid in demonstrating good performance results, and accountability to Parliament and the public for the government's regulatory activity. It provides the users, reviewers, internal auditors, and managers with information on how review can be used as a tool to help ensure that regulatory polices contribute to the public welfare of Canadians in a cost effective manner. In order to achieve this goal, the guide integrates the value of the review perspective in reforming existing or proposed regulatory programs. It presents a checklist of review principles that can be used as a method to identify appropriate regulatory or alternative approaches to program delivery. In this way, the guide establishes a foundation from which to discuss regulations and alternatives to regulations in relation to the complex obligations the federal government faces.

Leydesdorff, Loet & Gauthier, Élaine (1996). The evaluation of national performance in selected priority areas using scientometric methods. Research Policy, 25, 431-450.

How effectively can emerging science-based technologies be coupled to national R&D systems? Dutch and Canadian priority programs in biotechnology and advanced materials are analyzed in terms of differential increases in scientific output by using scientometric indicators and mappings. Methodological issues about using scientometric methods for science policy evaluations in the case of interdisciplinary and rapidly changing areas of 'techno-science' are discussed.
The paper is based on a comparative study of strategic research programs in Canada and the Netherlands.

Lipsett, Morley S., Holbrook, J. Adam, Lipsey, Richard G. &Wit, Robert W. de (1995). R&D and innovation at the firm level: improving the S&T policy information base. Research Evaluation, 5(2), 123-129.

Recent studies have indicated that the number of firms engaged in R&D has been significantly underestimated, which has hampered the evaluation of the impact of government S&T policies and programs on individual firms. The results of a study reported here suggest there is a need to supplement R&D data collected according to Frascati Manual standards to include information on 'part-time' R&D performers.

McDonald, Robert & Teather, George (1997). Science and Technology Evaluation Practices in the Government of Canada. In OECD (Ed.), Policy Evaluation in Innovation and Technology Towards Best Practices (Chapter 23). Paris: OECD.
http://www.oecd.org/dsti/sti/stat-ana/prod/evaluation.htm

McDonald, Robert & Teather, George (1997). Science and Technology Evaluation Practices in the Government of Canada. Ottawa: Industry Canada / National Research Council of Canada.
http://www.pmn.net/contributions/science/oecd.htm

Miller, Roger (1992). The influence of primary task on R&D laboratory evaluation: a comparative bibliometric analysis. R&D Management, 22(1), 3-20.

After reviewing the various approaches to R&D evaluation, the author concludes that a taxonomic approach combining organizational and bibliometric indicators offers a valid option to assess the quality of research.

Mothe, J. de la & Dufour, P. (Eds.). (1993). Science and Technology in Canada. Harlow, Essex: Longman Publishing. (ISBN 0582101069)

Muir, Langley R. & Williams, Douglas (1994). Management of R&D program evaluations: a case study of Canada's energy R&D program. Research Evaluation, 3(2), 97-106.

A large and complex evaluation of a Can$1.3 billion energy R&D program was recently carried out for the Canadian government. The challenges included a controversial and complex program, multiple stakeholders with very diverse interests, a large investment, an atmosphere of suspicion and mistrust and very tight deadlines. The methods used in the management of the study and which led to its successful completion are applicable to similar large-scale evaluation situations.

Smith, W. A. (1995). Evaluating research, technology and development in Canadian industry: Meeting the challenges of industrial innovation. Scientometrics, 34(3), 527-539.

Social Sciences and Humanities Research Council (1996). Strategic Grants Programs Review. Final Report. Ottawa: SSHRC.

Social Sciences and Humanities Research Council (1995). Evaluation of the General Research Grants Programme. Final Report. Ottawa: SSHRC.

Social Sciences and Humanities Research Council (1994). Post-doctoral Fellowships Evaluation Study. Ottawa: SSHRC.

Stoicheff, Boris P. (1999). Evaluation of Researchers, Research Proposals and Accomplishments. In Vaclav Paces, Ladislav Pivec & Albert H. Teich (eds.), Science Evaluatiion and Its Management (67-74). Ohmsha: IOS Press, NATO Science Series: Science & Technolgy Policy, vol. 28.

In spite of the ever changing environment for science in individual countries and globally, there is no better system for the evaluation of science at all levels than peer review. While modifications in carrying out peer review must be continually responsive to changes in resources and objectives, peer review by active scientists ensures excellence, creativity, and efficiency in research. Methods of evaluation of proposals and accomplishments are discussed with examples from Canada (e.g., The Natural Sciences and Engineering Research Council of Canada; The Canadian Institute for Advanced Research; Ontario Centres of Excellence).

Sylvain, C. (1993). Canadian research activity in aquaculture: A bibliometric analysis. Scientometrics, 27(3), 295-316.

Wilk, M. B. (1997). Are the costs and benefits of health research measurable? Ottawa, Ontario: Statistics Canada, Science and Technology Redesign Project, Research Papers No. 4.

Williams, Douglas & Rank, Dennis A. (1998). Measuring the economic benefits of research and development: the current state of the art. Research Evaluation, 7(1), 17-30.

The methodology for measuring the economic benefits of R&D has been considerably refined since the mid-1980s. Recently, increasing attention has been paid to the indirect benefits of R&D, in particular, the measurement of competency benefits. The paper shows that it is now possible to provide defensible estimates of both direct benefits (those arising from the use of the research results) and benefits which arise from the use of the competencies that are developed in the R&D process. A number of specific methodological advances are also discussed, such as the refinement of the notions of incrementality and attribution.

Hochschulevaluation Kanada

Association of Universities and Colleges of Canada (1995). A primer on performance indicators. Ottawa: AUCC.

Langford, Cooper H. (1999). The Evaluation of Research Done in Post-Secondary Institutions. Toronto, Ontario: Council of Ministers of Education of Canada. http://www.cmec.ca/postsec/index.stm

Maclean (1999). Universities Ranking. http://www.macleans.ca/pubdoc/1999/11/15/Universities1999/index.shtml

This site contains the latest results of Maclean's annual study of Canadian institutions of higher education. Rankings are provided for three groups of institutions: medical-doctoral (broad range of PhD programs and focus on research), comprehensive (significant research activity and wide range of undergraduate and graduate programs), and those focusing primarily on undergraduate education. The schools are evaluated in many areas, including faculty, classes, finances, libraries, reputation, and student body.

Mallea, John (1996). The evaluation of the higher education system in Canada. In Robert Cowen (Ed.). World Yearbook of Education 1996: The Evaluation of Higher Education Systems (51-59). London: Kogan Page.

Milot, Louise (1995). Relevance and Limitation of Periodic Programme Evaluation: The Case of Laval University. Higher Education Management, 7(1), 15-24.

An explanation of the periodic evaluation of Laval University (Canada) programs begins with an outline of the university's organizatinal structure and administrative bodies. The justification, objectives, and scope of the evaluation policy are then described, and some benefits and problems are discussed.

Snowdon, Ken (1994). The Use and Potential of Performance Indicators. Halifax, Nova Scotia: Canadian Institutional Researchers and Planners Association. http://www.usask.ca/cirpa/halifax94/potential/potential.html

Increasing interest in measures of accountability must be seen as an opportunity for post-secondary institutions to improve and strengthen the relationship with public sector and private sector constituents. This paper focuses on the policy and planning framework for the use of such indicators in a specific institution and points toward the use of indicators as one of the vehicles to address the issues surrounding accountability. Implications for institutional research are explored with an emphasis on improving organizational effectiveness.

Wolfe, David A. (1998). Quality and Accountability in PSE Research: The Measurement Challenge. Toronto, Ontario: Council of Ministers of Education of Canada. http://www.cmec.ca/postsec/index.stm

The paper highlights the complexity of the issues involved in using existing measures or developing new ones to evaluate the quality and effectiveness of the research effort at Canada's Post-secondary Education (PSE) institutions. It describes the three functions of the science system (knowledge production, knowledge transmission, knowledge transfer) relevant to the innovation process and reviews the existing state of available data sources that could be used to measure research effectiveness along these lines. It also explores some lines of recent academic research that indicate additional types of measures that might be used. New measures should be developed that more accurately capture the contribution of the research system to the innovation system.

2. USA

Forschungsevaluation

Abt Associates (1996). An Evaluation of the NSF Science and Technology Centers (STC) Program. Cambridge, MA.

Alston, Julian M., Norton, George W. & Pardey, Philip G. (1998). Science under Scarcity: Principles and Practice for Agricultural Research Evaluation and Priority Setting. New York: CAB International. (ISBN 0-85199-299-4)

Resources for agricultural science are scarce across the world. Yet even as resources are shrinking, agricultural science has expanded its inquiry into many new areas - such as environmental preservation, food quality, and rural development - without forsaking its more traditional concerns. I a time of tight government budgets, research administrators are faced with the need to provide strong evidence that costs are justified by benefits. Science under Scarcity is an invaluable guide to the theory and methods necessary for evaluating research in agriculture and for setting priorities for resource allocation. The book reviews, synthesizes, and extends such methods as economic surplus analysis, econometric techniques, mathematical programming procedures, and scoring models. The book was originally published in 1995 by Cornell University Press.

Averch, Harvey (1991). The practice of research evaluation in the United States. Research Evaluation, 1(3), 130-136.

The practice of research evaluation in the US is not very widespread, although demand is increasing. The last several years have seen significant improvements in our abilities to evaluate research. Evaluators need to persuade decision sponsors about the worth of investing modest sums in research and experiments on research evaluation.

Baldi, Stéphane (1998). Normative versus social constructivist processes in the allocation of citations: A network-analytic model. American Sociological Review, 63, 829-846.

The results of the study suggest that authors are likely to cite those articles most relevant to their work in terms of intellectual content, and seem little concerned with the characteristics of authors who write them.

Baldwin, Wendy & Seto, Belinda (1997). Peer: Review: Selecting the Best Science. Science and Engineering Ethics, 3(1), 11-17.

The major challenge facing today's biomedical researchers is the increasing competition for available funds. The competitive review process, through which the National Institutes of Health (NIH) awards grants, is built upon review by a committee of expert scientists. The NIH is firmly committed to ensuring that its peer review system is fair and objective.

Bird, Stephanie J. & Dustira, Alicia K. (Eds.). (1999). Scientific Misconduct (Special Issue). Science and Engineering Ethics, 5(2), 129-304.
http://www.opragen.co.uk/

Bozeman, Barry & Melkers, Julia (Eds.). (1993). Evaluating R&D Impacts: Methods and Practice. Norwell, MA: Kluwer.

The book is divided into two sections. The first section provides an introduction and details on several popular methodologies used in the evaluation of research and development activities (e. g., methods for evaluating the returns on R&D investments; the use of case studies in R&D impact evaluations; the use of bibliometric techniques in ex-post R&D evaluation; the use of co-word analysis for the evaluation of R&D; peer review approaches to R&D evaluations; patent analysis;

techniques and models from operations research applied to the evaluation of R&D projects). The second half of the book focuses on evaluation in practice and is comprised of several chapters offering the perspective of individuals in different types of organizations (e. g., evaluation of federal R&D; use of quantitative models to support research decision-making in business and government; evaluating the relevance of public laboratories‘ R&D; evaluation of strategic research programs; inventions program evaluation; evaluating the utility of individual projects and portfolios of projects). The book concludes with an annotated bibliography on evaluation of research, 1985-1990.

Branscomb, Lewis M. & Keller, James H. (Eds.). (1998). Investing in Innovation: Creating a Research and Innovation Policy that Works. Cambridge, MA: The MIT Press.

The second part of the book assesses seven specific technology programs promoted by the Clinton-Gore administration: (1) the Advanced Technology Program; (2) the Technology Reinvestment Project (Dual-Use Applications Program); (3) the Small Business Innovation Research Program; (4) the Technology Transfer Strategies at the National Institutes of Health; (5) the Manufacturing Extension Centers; (6) the Environmental Technology Initiative; and (7) Federal Energy Research.

Brown, E. A. (1996). Conforming the government R&D function with the requirements of the government performance and results act - Planning the unplannable? Measuring the unmeasurable? Scientometrics, 36(3), 445-470.

The Army Research Laboratory (ARL) was designed a Pilot Project for Performance Planning under the Government Performance and Results Act of 1993. Of the more than 80 such pilot projects government-wide, ARL was the only organization to represent the R&D community. As such, it was required to break new ground in both the planning and the evaluation of basic and applied research. The paper discusses the efforts made by ARL in both these areas, the insights drawn from these efforts, and the lessons learned.

Burnham, James B. (1997). Evaluating Industry/University Research Linkages. Research • Technology Management, 40(1), 52-55.

Joint industry-university research activity is an increasingly common form of conducting both basic and applied research.

Chubin, D. E. (1987a). Research Evaluation and the Generation of Big Science Policy. Knowledge, 9, 254-277.

Chubin, D. E. (1987b). Designing Research Program Evaluations: A Science Studies Approach. Science and Public Policy, 14, 82-90.

Collins, Eileen (1997). Performance Reporting in Federal Management Reform. Arlington, Virginia: National Science Foundation, Division of Science Resources Studies. http://www.nsf.gov/sbe/srs/perform/start.htm

Performance reporting in the Government Performance and Results Act of 1993 (GPRA) is part of a larger system to be adopted by each Federal agency in order to integrate planning, budgeting, management, and performance assessment.
Statements of concepts and principles from the Federal Accounting Standards Advisory Board (FASAB) and guidance from OMB emphasize the importance of measuring the results of government operations. All agree that it would be exceedingly convenient if a few comprehensive measures for each major Federal program could provide a complete and accurate charting of the program's results year after year. But reference is also occasionally made to the limitations of performance measurement and the need to supplement measures with other kinds of information in order to provide a complete and balanced picture.
Agencies and OMB are now in the process of determining the best balance between what can be measured and what cannot be measured and forging the most informative way to report results.

Meanwhile, auditors are concerned about how the proposed approaches to performance assessment can be audited. Agency staff and stakeholders need to be included in the developmental process. But few staff or stakeholders know the arcane details of performance reporting in the new management mandates.

Commission on Physical Sciences, Mathematics, and Applications (1995). Research Restructuring and Assessment. Washington, D.C.: National Academy Press.

Committee on Science, Engineering, and Public Policy of the National Academy of Sciences, National Academy of Engineering, Institute of Medicine (1999). Evaluating Federal Research Programs - Research and the Government Performance and Results Act. Washington, DC: National Academy Press.
http://www.nap.edu/books/0309064309/html/R1.html

The Government Performance and Results Act (GPRA), passed by Congress in 1993, requires that federal agencies write five-year strategic plans with annual performance goals and produce an annual report that demonstrates whether the goals have been met. The first performance reports are due in March 2000. Measuring the performance of basic research is particularly challenging because major breakthroughs can be unpredictable and difficult to assess in the short term. This book recommends that federal agencies use an "expert review" method to examine the quality of research they support, the relevance of that research to their mission, and whether the research is at the international forefront of scientific and technological knowledge. It also addresses the issues of matching evaluation measurements to the character of the research performed, improving coordination among agencies when research is in the same field, and including a human resource development component in GPRA strategic and performance plans.

Committee on Science, Engineering, and Public Policy of the National Academy of Sciences, National Academy of Engineering, Institute of Medicine (2000). Experiments in International Benchmarking of US Research Fields. Washington, DC: National Academy Press.
http://books.nap.edu/catalog/9784.html

To assess the feasibility and utility of international benchmarking, COSEPUP carried out a set of experiments in three fields: mathematics, immunology, and materials science and engineering. The results of the experiments suggests that research leadership status by field can be assessed in a timely fashion at reasonable cost.

Committee on Science, Engineering, and Public Policy of the National Academy of Sciences, National Academy of Engineering, Institute of Medicine (1999). International Benchmarking of US Immunology Research. Washington, DC: National Academy Press.
http://books.nap.edu/catalog/9444.html

Committee on Science, Engineering, and Public Policy of the National Academy of Sciences, National Academy of Engineering, Institute of Medicine (1998). International Benchmarking of US Materials Science and Engineering Research. Washington, DC: National Academy Press.

Committee on Science, Engineering, and Public Policy of the National Academy of Sciences, National Academy of Engineering, Institute of Medicine (1997). International Benchmarking of US Mathematics Research. Washington, DC: National Academy Press.

Committee on Science, Engineering, and Public Policy of the National Academy of Sciences, National Academy of Engineering, Institute of Medicine (1996). An Assessment of the National Science Foundation's Science and Technology Centers Program. Washington, DC: National Academy Press.

NSF requested COSEPUP to conduct a study of the STC program. COSEPUP appointed a panel to carry out this study. The panel was to review and interpret the data gathered by an outside contractor to NSF (Abt Associates), reach conclusions regarding the progress of the STCs program toward its goals, and make recommendations concerning NSF's future use of the STC mode of support. The use of Abt Associates as well as the Academy was an experimental effort to have a contractor develop some of the empirical data desired by NSF. The COSEPUP panel does not view the experiment as successful.

Committee on Science, Engineering, and Public Policy of the National Academy of Sciences, National Academy of Engineering, Institute of Medicine (1994). Major Award Decisionmaking at the National Science Foundation. Washington, DC: National Academy Press.

Committee on the NIH Research Priority-Setting Process of the Institute of Medicine (1998). Scientific Opportunities and Public Needs: Improving Priority Setting and Public Input at the National Institutes of Health. Washington, DC: National Academy Press. http://books.nap.edu/catalog/6225.html

The National Institutes of Health (NIH) is the single largest funder of health research in the United States, and research it has supported has been pivotal to the explosion of biomedical knowledge over the past century. As NIH's success has grown, so has pressure from advocacy groups and other members of the public to devote more spending to their health concerns. In response to a request from Congress, this IOM study reviewed NIH's research priority-setting process and made recommendations for possible improvement. The committee considered the:
· Factors and criteria used by NIH to make funding allocations.
· Process by which the funding decisions are made.
· Mechanisms for public input.
· Impact of congressional statutory directives on funding decisions.
Among other recommendations, the book recommends that NIH seek broader public input on decisions about how to spend its nearly $14 billion budget; it also urged the agency to create new Offices of Public Liaison in the Office of the Director and in each of the 21 research institutes to allow interested people to formally take part in the process.

Coursey, Bert M. & Link, Albert N. (1998). Evaluating technology-based public institutions: the case of radiopharmaceutical standards research at the National Institute of Standards and Technology. Research Evaluation, 7(3), 147-157.

The purpose of the paper is to illustrate, through one case study, the current state of program evaluation in the reseach laboratories at the US National Institute of Standards and Technology (NIST). The history of NIST's radiopharmaceutical standards research program is discussed, the methodology for data collection and analysis is detailed, and the NIST management's use of the findings from the case study is described, in an effort to move toward generalizations about best practices in program evaluation applicable to technology-based public institutions.

Cozzens, Susan E. (1996). Quality of Life Returns from Basic Research. In: Bruce L. R. Smith & Claude E. Barfield (Eds.), Technology, R&D, and the Economy (184-209). Washington, DC: The Brookings Institution and American Enterprise Institute.

What has basic research contributed to the quality of life? What criteria should be used to judge whether current efforts are succeeding? What goals can be set and how can research be managed to achieve the returns in quality of life desired now and for future generations? A great deal of work needs to be done to increase the effectiveness of programs with regard to goals that will affect the quality of life. The challenges involve (1.) putting the researcher back at the center of research evaluation and planning concepts. The focus too long and too exclusively has been on the knowledge products of research. (2.) Putting the public back in evaluation and planning processes. Public involvement in research management has been seen for too long as a threat to autonomy and a form of political control. Public involvement in program evaluation and planning my be the only

route, under current circumstances, to leaving researchers in control of their research topics and processes.

Cozzens, Susan E. (1995). U. S. Research Assessment: Recent Developments. Scientometrics, 34(3), 351-362.

Over the last decade, ex post research assessment at the program level in the U. S. has seemed much less active than the equivalent activities in Europe. This seeming lull was the result of a decline in program evaluation activity across the U. S. government in the 1980s, which slowed the rate of formal evaluations. Program review activities within agencies, however, were common, especially at such mission-oriented research supporting organizations as the Department of Energy and the Office of Naval Research. Review processes at these agencies relied primarily on expert assessment, sometimes at the project level, supplemented by user inputs. Quantitative performance measures were seldom used. That situation is about to change. In 1993, Congress passed the Government Performance and Results Act, which requires all agencies including those support research to set quantitative performance targets and report annually on their progress toward them. Agencies with clear technological goals are rapidly developing sets of indicators for this use, including peer assessments, bibliometric measures including patents, and customers satisfaction ratings. But fundamental research agencies do not find such measures satisfactory, and are just beginning to develop alternative ones.

Cozzens, Susan E. (1987). Expert Review in Evaluating Programs. Science and Public Policy, 14, 71-81.

Cozzens, Susan E. & Melkers, Julia E. (1997). Use and Usefulness of Performance Measurement in State Science and Technology Programs. Policy Studies Journal, 25(3), 425-435.

Over 40% of state science and technology programs are evaluating their results in some way. Programs have adopted performance measurement systems both for their own information purposes and to justify their activities to external audiences, usually legislatures. Performance measurement activities may have contributed to the evolution of these programs toward economic objectives, which can be measured in business activity. The most common performance measures are job creation and retention.

Cozzens, Susan E. & Melkers, Julia E. (1994-1996). Science and Technology-Based Economic Development Programs in the States: A Study of Evaluation Efforts (Project Summary). Funded by the National Science Foundation, Grant # SBR-9422433. http://www.gsu.edu/~padjem/projects.html

This research examined evaluation activities of science and technology-based (S&T) programs at the state level in the United States. S&T-based programs are those that stimulate the science base or technology development in a geographic area with the purpose of attracting or stimulating industrial investment and development there. The goal of this project was to survey the existing approaches and methods of evaluation in order to improve state-level practice. The project involved a series of case studies and a mail survey of state S&T programs. Findings show that performance measurement activities are common in close to half the state programs. For some states, measurement is required as part of state performance-based budgeting. Other programs, however, collect performance data for program improvement reasons although they are not required to do so. The nature of the data collected varies from state to state. However, most programs are pressured, especially by legislators, to report job creation information. Overwhelmingly, the influence of the organization's director plays a critical role in the commitment to evaluation and performance measurement activities. Further, this influence is important in the type and extent of the actual use of the performance data.

Cronin, Blaise & Overfelt, Kara (1994). Citation-Based Auditing of Academic Performance. Journal of the American Society for Information Science, 45(2), 61-72.

The use of citation data in evaluating the research performance of academic programs and individual faculty members is explored in the context of a 10-year analysis of a single academic unit. The study controls for possible accounting bias by comparing results obtained using three differing approaches to allocation citation credit: straight, whole, and adjusted counts. Citation scores are correlated with salary, time-in-field, and gender. The results of the study raise serious questions about the validity of research rankings derived from subjective perception studies.

Cullars, John (1992). Citation characteristics of monographs in the fine arts. Library Quarterly, 62(3), 325-342.

Diodato, Virgil (1994). Dictionary of Bibliometrics. New York: Haworth Press.

Bibliometrics is a field that uses mathematics and statistical techniques to study publishing and communication patterns in the distribution of information. The dictionary explains some 225 terms used in bibliometrics. Its first purpose is to give the reader nontechnical definitions of bibliometric concepts (e.g., bibliographic coupling, cocitation). The second purpose of the dictionary is to suggest sources where the reader can find more information about the defined term.

Eldon, R. E. & Devine, C. M. (1985). Government's Research and Evaluation Gap. Public Relations Review, 11, 47-56.

A survey of 131 top federal information officers shows the gap between the rhetoric about the importance of research and evaluation in decision making and their actual use. The results of this survey are consistent with the view that research and evaluation are more important symbolically than substantively and with the view that government agencies and decision makers rationally demand much more systematic information than they want to use.

Ellis, Lynn (1997). Evaluation of R&D Processes: Effectiveness Through Measurements. Norwood, MA: Artech House.

The book emphasizes a quantitative approach to evaluation based on the philosophy that „what gets measured, gets done; so be sure you measure what you want!“. In recent years, the metrics chosen by industrial companies have branched out from being purely economic, financial, and accounting also to include nonfinancial metrics such as measures of customer satisfaction, timeliness, and quality, to name a few.

Endres, Al (1997). Improving R&D Performance - the Juran Way (Chapter 3: Measuring R&D Quality, Chapter 4: Assessing R&D Quality Status). New York: Wiley. ISBN 0-471-16370-8

The author describes the application of Total Quality Management (TQM) in Research and Development.

Feller, I. (1988). Evaluating State Advanced Technology Programs. Evaluation Review, 12, 232-252.

Frankel, Mark S. & Cave, Jane (Eds.) (1997). Evaluating Science and Scientists: An East-West Dialogue on Research Evaluation in Post-Communist Europe. Budapest: Central European University Press. (ISBN 1-85866-079-3)

The volume examines efforts to reform research funding and evaluation in Eastern Europe, and provides an overview of Western experience and scholarly research related to the use of peer review and quantitative evaluation techniques. The authors examine the political and economic contexts in which resources for science are allocated, the underlying assumptions of peer review and the use of quantitative indicators, the reliability of those methods, their weaknesses and possible remedies, and their consequences for science and other social institutions. Also presented are new ideas for practical steps to minimize the technical, administrative and ethical problems raised by the use of peer review and quantitative techniques in evaluating science and scientists.

Garfield, Eugene (1998). From Citation Indexes to Informetrics: Is the Tail Now Wagging the Dog? Libri, 48, 67-80.

The article provides a synoptic review and history of citation indexes and their evolution into research evaluation tools including a discussion of the use of bibliometric data for evaluating U. S. institutions (academic departments) by the National Research Council (NCR).

Garfield, Eugene (1996). When to cite. Library Quarterly, 66(4), 449-458.

Garfield, Eugene & Welljams-Dorof, Alfred (1992). Citation data: their use as quantitative indicators for science and technology evaluation and policy-making. Science and Public Policy, 19(5), 321-327.

Geisler, Eliezer (1996). Integrated Figure of Merit of Public Sector Research Evaluation. Scientometrics, 36(3), 379-395.

The Integrated Figure of Merit (IFM) is a combination of a cost model of research and a performance model. IFM incorporates elements of quantitative and judgmental indicators of outputs, and allows for comparisons of the same organization over time, as well as among different organizations in the same time frame. This is a central requirement of a useful evaluation system.

Geisler, Eliezer (1994). Key Output Indicators in Performance Evaluation of Research and Development Organizations. Technological Forecasting and Social Change, 47(2), 189-203.

General Accounting Office (1999). Federal Research: Peer Review Practices at Federal Science Agencies Vary. GAO/RCED-99-99.
http://www.gao.gov/AIndexFY99/abstracts/rc99099.htm

The federal government will invest $80 billion in fiscal year 1999 on research and development done by government scientists and through grants, contracts, or other agreements with universities, corporations, small businesses, and other members of the research community. GAO was asked to study the peer review and other quality assurance processes that federal agencies use in conducting scientific research and development. GAO reviewed 12 federal agencies to (1) define what is meant by peer review, (2) describe the federal government's peer review policy, (3) describe the peer review practices of the agencies that conduct scientific research, (4) describe other agency quality assurance reviews, and (5) identify which research is not subject to review. GAO found that there is no written governmentwide definition of peer review and that each of the 12 agencies had various policies, orders, or other internal guidance on the conduct of peer reviews.

General Accounting Office (1994). Peer Review: Reforms needed to ensure fairness in federal agency grant selection. Washington, DC: U.S. Government Printing Office. GAO/PEMD-94-1.

General Accounting Office (1997). Measuring Performance: Strengths and Limitations of Research Indicators. Washington, DC: United States General Accounting Office. GAO/RCED-97-91.
http://www.gao.gov/cgi-bin/titlesearch.pl

General Accounting Office (1997). Measuring Performance: Challenges in Evaluating Research and Development. Washington, DC: United States General Accounting Office. GAO/T-RCED-97-130.
http://www.gao.gov/cgi-bin/titlesearch.pl

General Accounting Office (1996). Measuring Performance: The Advanced Technology Program and Private-Sector Funding. Washington, DC: United States General Accounting

Office. GAO/RCED-96-47.
http://www.gao.gov/cgi-bin/titlesearch.pl

General Accounting Office (1997). Results Act: Observations on Federal Science Agencies. Washington, DC: United States General Accounting Office. GAO/RCED-97-220
http://www.gao.gov/cgi-bin/titlesearch.pl

GAO Reports on the Government Performance and Results Act:
http://www.gao.gov/new.items/gpra/gpra.htm

The Government Performance and Results Act of 1993 seeks to shift the focus of government decisionmaking and accountability away from a preoccupation with the activities that are undertaken - such as grants dispensed or inspections made - to a focus on the results of those activities, such as real gains in employability, safety, responsiveness, or program quality. Under the Act, agencies are to develop multiyear strategic plans, annual performance plans, and annual performance reports.

Agencies' Strategic Plans and Performance Plans:

Department of Agriculture
Strategic Plan: http://www.usda.gov/ocfo/strat/
Performance Plan: http://www.usda.gov/ocfo/annlplan/index.html

Department of Defense
Strategic Plan: http://www.defenselink.mil/pubs/qdr/
Performance Plan: http://www.dtic.mil/execsec/adr98/apdx_j.html

Department of Energy
Strategic Plan: http://www.doe.gov/policy/doeplan.html
Performance Plan: http://www.doe.gov/policy/sol98/index.htm

Department of Health and Human Services
Strategic Plan: http://aspe.os.dhhs.gov/hhsplan/intro.htm
Performance Plan: http://www.hhs.gov/progorg/asmb/budget/fy99budget

Department of Transportation
Strategic Plan: http://www.dot.gov/hot/dotplan.html
Performance Plan: http://ostpxweb.dot.gov/budget/perfp99.htm

Environmental Protection Agency
Strategic Plan: http://www.epa.gov/ocfopage/plantoc.htm
Performance Plan: http://www.epa.gov/ocfo/99budget/1999bib.htm

National Aeronautics and Space Administration
Strategic Plan: http://www.hq.nasa.gov/office/nsp/

National Oceanic and Atmospheric Administration
Strategic Plan: http://www.noaa.gov/str-plan/
Performance Plan: http://www.doc.gov/bmi/budget/strtgc/strtone.htm

National Institute of Standards and Technology
Strategic Plan: http://www.doc.gov/bmi

Performance Plan: http://www.doc.gov/bmi/budget/strtgc/strtone.htm

National Science Foundation
Strategic Plan: http://www.nsf.gov/od/gpraplan/gpraplan.htm
Performance Plan: http://www.nsf.gov/pubs/1998/nsf99gprapp/start.htm

Hackett, Edward J. (1997). Peer Review in Science and Science Policy. In Mark S. Frankel & Jane Cave (Eds.), Evaluating Science and Scientists: An East-West Dialogue on Research Evaluation in Post-Communist Europe (51-60). Budapest: Central European University Press.

The author argues that peer review is much more than a set of practices and principles for allocating rewards and resources. Too often, he says, analysts focus exclusively on the mechanics of peer review and produce erroneous diagnoses of perceived malfunctions in the system because they fail to understand the multiple functions of, and demands made on, peer review.

Hauser, John, R. & Zettelmeyer, Florian (1996). Metrics to Evaluate R, D & E. Cambridge, MA: MIT, Sloan School of Management, International Center for Research on the Management of Technology.
WP # 156-96

Henderson, Rebecca, Jaffe, Adam B. & Trajtenberg, Manuel (1998). Universities as a source of commercial technology: a detailed analysis of university patenting, 1965-1988. The Review of Economics and Stastistics, 119-127.

The paper explores the recent explosion in university patenting as a source of insight into the changing relationship between the university and the private sector. Before the mid-1980s, university patents were more highly cited, and were cited by more diverse patents, than a random sample of all patents. More recently several significant shifts in university patenting behavior have led to the disappearance of this difference. Thus our results suggest that between 1965 and 1988 the rate of increase of important patents from universities was much less than their overall rate of increase in patenting.

Hertzfeld, Henry (1998). Research into Economic Methodology and Models for Assessing Impacts of NASA Life Science Research Investments. Washington, DC: The George Washington University, The Elliott School of International Affairs, Space Policy Institute.

A survey of forty-one companies that reported prior commercial success in transforming NASA R&D investments in the life sciences into marketable goods and services was conducted in late 1997 by the Space Policy Institute, George Washington University. Fifteen of these firms provided useful data for this study. These firms alone have cumulatively contributed over $1.5 billion in value added to the economy over the past twenty-five years. The cumulative NASA R&D investment in the technologies represented by the products of these firms was approximately $64 million. An additional $200 million in private R&D from those companies was stimulated by the NASA investment. This additional R&D was necessary for the production, development, and marketing of the commercial products and represents the positive leverage of NASA life sciences investments. These are conservative estimates because they only measure the impact of NASA R&D on the companies that produce and market the products. The results from personal interviews conducted for this study also show that there are very large benefits that accrue to the purchasers and users of the life sciences products produced and sold by these companies. These social benefits range from cost savings through the use of more efficient medical and research equipment to non-quantifiable benefits such as the substitution of non-invasive procedures for surgery. These societal and downstream impacts and benefits are documented and described in this study. http://www.gwu.edu/~spi/research.html

Hertzfeld, Henry (1998). Measuring the returns to NASA life sciences research and development. Washington, DC: The George Washington University, The Elliott School of International Affairs, Space Policy Institute.

Jaffe, Adam B. & Trajtenberg, Manuel (1999). International knowledge flows: evidence from patent citations. Econ. Innov. New Techn., 8, 105-136.

Jaffe, Adam B., Fogarty, Michael S. & Banks, Bruce A. (1997). Evidence from patents and patent citations on the impact of NASA and other federal labs on commercial innovation. Cambridge, MA: National Bureau of Economic Research. Working Paper 6044. http://www.nber.org

Kleinknecht, Alfred & Bain, Donald (Eds.). (1993). New Concepts in Innovation Output Measurement. New York: St. Martin's Press.

Kostoff, Ronald N. (1998). Research Evaluation Documents. Arlington, VA: Office of Naval Research.
http://www.dtic.mil/dtic/kostoff/index.html

Research Evaluation Documents:
This web site contains five documents focused on research evaluation and impact assessment, authored by Dr. Ronald N. Kostoff, Office of Naval Research.
The newest monograph entitled „SCIENCE AND TECHNOLOGY INNOVATION“ describes two novel complementary approaches for systematically enhancing the process of innovation and discovery. One approach is workshop-based and the other approach is literature-based. Both approaches have the common feature of exploring knowledge fromvery disparate technical disciplines and technologies, and transferring insights and understanding from one or more disparate technical areas to another. While either approach can be performed independently to enable innovation and discovery, it is highly recommended that the approaches be combined into a single process. This integrated approach utilizes the strengths of each component technique to provide a synergy which can lead more efficiently to innovation than the sum of the two approaches performed separately. It has the potential to be a major breakthrough for the systematic promotion of innovation and discovery.
A monograph entitled „SCIENCE AND TECHNOLOGY METRICS“ was added to the web site in April 1998. This document describes: a) why S&T assessment and evaluation have become important; b) why metrics have become important for quality S&T evaluation; c) what types of metrics are available for S&T evaluation, and d) how metrics have been and can be applied to prospective and retrospective S&T assessment and evaluation. Many case studies of metrics applications are summarized. The monograph discusses how metrics can be integrated with other evaluation tools to address the requirements of the Government Performance and Results Act of 1993. This comprehensive document is self-contained, with 14 Appendices, and can serve as an information resource with over 5600 text and suggested reading references.
The article on peer review covers issues and concerns, principles for high quality reviews, federal agency practices, and contains an extensive bibliography. A brief summary of the article, in the context of peer review as the appropriate GPRA metric for basic research, has been published in Science Magazine (Policy Forum, 1 August 1997). The article on roadmaps identifies: roadmap concerns and issues; how roadmaps can support quality peer review; characteristics of quality roadmaps; selected examples of different roadmap types; and contains an extensive bibliography. The Handbook of Research Impact Assessment is an extensive compendium of methods for evaluating research and its impacts, and contains a total of about 4,000 references.

Kostoff, Ronald N. (1998). Research Program Peer Review: Principles, Practices, Protocols. Arlington, VA: Office of Naval Research.
http://www.dtic.mil/dtic/kostoff/Peerweb1index.html

The principles, practices, and protocols of research program peer review are described. While the principles are fundamentally generic, and apply to peer review across the full spectrum of performing institutions, as well as manuscript/ proposal/ program peer review, the focus of this paper is peer review of proposed and ongoing research programs in federal agencies.
Following the self-contained Executive Summary of factors for high-quality peer reviews, the paper addresses potential implications of the implementation of the Government Performance and Results Act of 1993 on federal agency research program peer review practices. Then, the paper describes strengths and weaknesses of major peer review components and issues including: Objectives and Purposes of Peer Review; Quality of Peer Review; Impact of Peer Review Manager on Quality; Selection of Peer Reviewers; Selection of Evaluation Criteria; Secrecy (Reviewer and Performer Anonymity); Objectivity/ Bias/ Fairness of Peer Review; Normalization of Peer Review Panels; Repeatability/ Reliability of Peer Review; Effectiveness/ Predictability of Peer Review; Costs of Performing a Peer Review; Ethical Issues in Peer Review; and Alternatives to Peer Review.
The paper then presents different federal agency peer review practices, and sample protocols and processes for conducting a successful research program peer review. Some peer review variants, such as the Science Court, are described, and research requirements to improve peer review are discussed. The final section is an extensive bibliography of over 1500 references which includes not only text references but related references for further reading as well.

Kostoff, Ronald N. (1998). The use and misuse of citation analysis in research evaluation. Scientometrics, 43(1), 27-43.

Kostoff, Ronald N. (1997). The Principles and Practices of Peer Review. Science and Engineering Ethics, 3(1), 19-34.

The principles and practices of research peer review are described. While the principles are fundamentally generic and apply to peer review across the full spectrum of performing institutions as well as to manuscript/proposal/program peer review, the focus of this paper is peer review of proposed and ongoing programs in federal agencies. The paper describes desireable characteristics and important intangible factors in successful peer review. Also presented is a heuristic protocol for the conduct of successful peer review research evaluations and impact assessments. Problems with peer review are then outlined, followed by examples of peer review of proposed and existing programs in selected federal agencies. Some peer review variants, such as the Science Court, are described, and then research requirements to improve peer review are discussed.

Kostoff, Ronald N. (1996). Performance measures for government-sponsored research: Overview and background. Scientometrics, 36(3), 281-292.

Kostoff, Ronald N. (1995). Federal research impact assessment: Axioms, approaches, applications. Scientometrics, 34(2), 163-206.

Kostoff, Ronald N. (1994). Federal Research Impact Assessment: State-of-the-Art. Journal of the American Society for Information Science, 45(6), 428-440.

Kroll, Peter, Ault, Grace & Narin, Francis (1998). Tracing the Influence of Basic Scientific Research on Biotechnology Patents. Patent World, 100, 38-46.

LaFollette (1994). Measuring Equity: The U.S. General Accounting Office Study of Peer Review. Science Communication, 16(2), 211-220.

Link, Albert N. (1998). Public Accountability: Evaluating Technology-Based Institutions. Norwell, MA: Kluwer.

Link, Albert N. (1996). Evaluating Public Sector Research and Development. Westport, CT: Praeger. (ISBN 0-275-95368-8)

The book presents the findings from seven in-depth case studies of R&D programs being conducted at the National Institute of Standards and Technology (NIST): Advanced Technology

Program; Real-Time Control System Architecture; Conformance Test Program for SQL; ISDN Technology; Power and Energy Calibration Services; Electromigration Characterization; and Optical Fiber Standards. These evaluation studies, undertaken over the past five years, illustrate through case analyses that information can systematically be gathered and preliminary inferences can be made about the relative benefits of public R&D programs.

Link, Albert N. (1996). Economic performance measures for evaluating government-sponsored research. Scientometrics, 36(3), 325-342.

MacRoberts, M. H. & MacRoberts, B. R. (1996). Problems with citation analysis. Scientometrics, 36(3), 435-444.

Melkers, Julia (1993). Bibliometrics as a tool for analysis of R&D impacts. In: Barry Bozeman & Julia Melkers (Eds.). Evaluating R&D Impacts: Methods and Practice (43-61). Norwell, MA: Kluwer.

Bibliometrics is a retrospective evaluation approach, useful for ex-post evaluations. However, there are several points to consider when applying bibliometric analysis to R&D evaluation: (1) use it on the appropriate unit of analysis; (2) use it only as a partial indicator; (3) use it for projects that have a strong science quality; (4) do not try to account for quality with bibliometrics; and (5) provide information about what data indicate and what they do not when presenting evaluation results.

Melkers, Julia & Roessner, David (1997). Politics and the political setting as an influence on evaluation activities: national research and technology policy programs in the United States and Canada. Evaluation and Program Planning, 20(1), 57-75.

Program evaluation activities in the United States, after a long history of decentralized, uncoordinated activity, have taken a new turn. It is called performance measurement. A portion of this history, and the recent changes, may be explained by characteristics of the political environment. In contrast, Canada has a long history of centralized, coordinated evaluation of its federal programs. In the paper, the authors identify particular attributes of the Canadian and the U.S. political systems that are related to each nation's respective evaluation system. Specifically, they address the following factors differentiating the evaluation experience in both countries as being a function of (1) the level of centralization; (2) legislative history; (3) legislative precedent / guidelines for evaluation; and (4) political support for evaluation in terms of infrastructure, rules, and requirements. Using case studies of research and development program evaluations in both countries, the authors examine the forces at work that shape the design and implementation of evaluation programs. The paper concludes with a discussion of the implications of recent changes in both systems for their respective evaluation systems and processes.

Mullins, N. C. (1987). Evaluating research programs: Measurement and data sources. Science and Public Policy, 14(2), 91-98.

Narin, Francis (1995). Patents indicators for the evaluation of industrial research output. Scientometrics, 34(3), 489-496.

Narin, Francis (1994). Patent Bibliometrics. Scientometrics, 30(1), 147-155.

Narin, Francis (1987). Bibliometric Techniques in the Evaluation of Research Programs. Science and Public Policy, 14(2), 99-106.

Narin, Francis & Breitzman, Anthony (1995). Inventive productivity. Research Policy, 24, 507-519.

An investigation of the number of patents per inventor was carried out for four companies, two American and two Japanese, in the area of semiconductors. For all four cases a Lotka-like distribution was found, with a relatively small number of highly productive inventors and a large number of inventors with their names on only one, and a general factor of ten difference in productivity between the most- and the least-productive inventors.

Narin, Francis & Hamilton, Kimberly S. (1996). Bibliometric performance measures. Scientometrics, 36(3), 293-310.

Three different types of bibliometrics - literature bibliometrics, patent bibliometrics, and linkage bibliometrics can all be used to address various Government Performance and Results questions.

Narin, Francis, Hamilton, Kimberly S. & Olivastro, Dominic (1997). The increasing linkage between U.S. technology and public science. Research Policy, 26, 317-330.

Narin, Francis, Hamilton, Kimberly S. & Olivastro, Dominic (1995). Linkage between agency-supported research and patented industrial technology. Research Evaluation, 5(3), 183-187.

CHI Research's early work on the citation linkage between patented technology in the USA, and the underlying research science base using 1987/88 US patents has been massively expanded to include citations from 1993/94 US patents, and an analysis of the cited US papers and the agencies supporting them. There is a very strong within-country component to the linkage: inventors in the US system cite their own country's papers approximately three times as often as would be expected, when adjusted for the size of the country's science. The linkage is strongest in the highly scientific areas of technology, and is quite subject specific. Over the six years separating the studies, there has been a remarkable three-fold increase in linkage. A large fraction of these papers cited in patents originate in the US university system, and are supported by US research support agencies.

Narin, Francis, Olivastro, Dominic & Stevens, Kimberly, A. (1994). Bibliometrics: Theory, Practice and Problems. Evaluation Review, 18(1), 65-76.

This article presents the theory behind modern evaluative bibliometric techniques at three levels. Policy applications, which characterizes the scientific and technological output of nations or regions; strategic analyses, which deals with articles and patents at the level of a university or company; and tactical analyses, which addresses questions concerning a single subject. The article explains the newer techniques that have been developed at each level, as well as the more important limitations.
Further CHI's Publications:
http://www.chiresearch.com/fullpubs.htm

National Academy of Sciences (1996). Beyond Discovery: The Path from Research to Human Benefit. Washington, D. C.: National Academy Press.
http://www4.nas.edu/beyond/beyonddiscovery.nsf

„Beyond Discovery: The Path from Research to Human Benefit“ is a project of the National Academy of Sciences. It is a series of case studies that identify and trace origins of important recent technological and medical advances. Each case study reveals the crucial role played by basic science, the applications of which could not have been anticipated at the time the original research was conducted.

National Science and Technology Council (1996). Assessing Fundamental Science. Washington, DC: National Science and Technology Council.
http://www.nsf.gov/sbe/srs/ostp/assess/start.htm

The 1993 Government Performance and Results Act (GPRA) emerged from a bipartisan effort to improve accountability, productivity, and effectiveness of Federal programs through strategic planning, goal setting, and performance assessment. The annual assessments will serve the dual purpose of guiding subsequent planning decisions and of communicating program outcomes and impacts to the public. This document, developed under the auspices of the Committee on Fundamental Science of the National Science and Technology Council (NSTC), serves to establish

a broad framework for GPRA implementation in assessment of fundamental science programs. GPRA anticipates the need for flexibility in designing the planning and assessment methodology appropriate to the great variety of Federal programs. Clearly, all programs are intended to contribute ultimately to over-arching national goals, such as national security, quality of life, and economic prosperity. Nevertheless, the manner in which specific programs do so is dramatically different. For example, procurement of military hardware contributes rather directly to the national security goal, whereas an undergirding activity, such as fundamental scientific research, contributes broadly to national goals over a very long time period. Assessment techniques are in relatively early stages of development in all areas and are only in their infancy for areas such as fundamental science. GPRA anticipates the need for time and experimentation in developing assessment techniques by building in a phase-in period of several years for implementation. The central issue in assessing fundamental science lies in defining the goal against which progress is measured. The Administration's science policy statement, Science in the National Interest, establishes that goal as leadership across the frontiers of scientific knowledge. This is the critical measure for assuring that American scientists are expanding the knowledge base at the leading edge. We stress that leadership evaluation does not entail simplistic numerical ranking of national programs. Our national interest in leadership rests in having our research and educational programs perform at the cutting edge - sometimes in competition, but often in explicit collaboration, with scientists from other nations.

National Science Board (1998). Science & Engineering Indicators - 1998 (Chapter 5-37 to 5-53: Outputs of Scientific and Engineering Research). Arlington, VA: National Science Foundation. (NSB-98-1)

National Science Foundation (1997). GPRA Strategic Plan for FY 1997-FY 2003. Arlington, Viginia 22230.
http://www.nsf.gov/od/gpraplan/gpraplan.htm

In February, 1995, the National Science Foundation (NSF) and the National Science Board published „NSF in a Changing World“, a strategic plan designed to guide NSF for five to ten years. Since its publication, it has served as a touchstone for all our activities, providing an overarching sense of purpose and direction.

The strategic plan at hand provides an operational implementation of „NSF in a Changing World“, in compliance with the Government Performance and Results Act (GPRA). It provides concrete outcome goals that are tied to the results of NSF's grants for research and education in science and engineering and also addresses our goals for excellence in managing the agency. It describes investment strategies and actions we intend to take to implement those strategies. It illustrates by example some of the many activities already underway that embody the strategies. It tells you much about what we intend to do, but, for reasons of brevity, it says very little about the uniqueness and importance of NSF's mission and the rationale for the approach we have chosen to follow. I believe the rationale is important in setting the proper context for the elements of this plan.

The organization of this GPRA strategic plan presents a set of key investment strategies for each of NSF's programmatic outcome goals. This makes the plan easier to read, but obscures the fact that there are a small number of quite general strategies that have impact across and, thus, create synergy among the outcome goals. These include NSF's commitment to: (1) using competitive merit review with peer evaluation to identify the most promising ideas from the strongest researchers and educators; (2) integrating research and education to strengthen both; (3) working in partnership with the science and engineering communities and potential users of the results of NSF investment in order to identify areas of emerging opportunity; and (4) assuring that both NSF and the research and education communities reap optimal benefit from the revolution in information, communications, and computing technologies.

This is the first stage of NSF's compliance with the Government Performance and Results Act. Annual performance plans and reports will come next.

National Science Foundation (1998). FY 1999 GPRA Performance Plan (revised January 1999). Arlington, Viginia 22230.
http://www.nsf.gov/cgi-bin/getpub?nsf99gprappr

This is NSF's first GPRA performance plan. It is based on NSF's GPRA strategic plan, submitted to Congress in September, 1997. The mission, outcome goals, and critical factors for success from that plan are outlined.

National Science Foundation (1999). FY 2000 GPRA Performance Plan. Arlington, Viginia 22230.
http://www.google.com/search?q=National+Science+Foundation+FY+2000+GPRA+Performance+Plan

This is NSF's second GPRA performance plan. It is based on NSF's GPRA strategic plan, submitted to Congress in September, 1997, and the FY 1999 performance plan. The mission, outcome goals, and critical factors for success from the strategic plan are outlined.
For additional information on NSF's GPRA, please visit the GPRA website at http://www.nsf.gov/od/gpra/start.htm

Nelson, Richard R. (1996). The sources of economic growth. Boston, MA: Harvard University Press. ISBN 0674821459

Drawing on a deep knowledge of economic and technological history as well as on the tools of economic analysis, the author describes the intimate connections among public policies, science-based universities, and the growth of technology. He compares national innovation systems, and explores both the rise of the United States as the world's premier technological power during the first two-thirds of the twentieth century and the diminishing of that lead as other countries have largely caught up.

Porter, Alan L., Roessner, J. David, Newman, Nils & Cauffel, David (1996). Indicators of high technology competitiveness of 28 countries. International Journal of Technology Management, 12(1), 1-32.

Radin, Beryl A. (1998). The Government Performance and Results Act (GPRA): Hydra-Headed Monster or Flexible Management Tool? Public Administration Review, 58(4), 307-316.

Roessner, David (1999). New Approaches to Evaluating Research Programmes for Management Purposes. In Vaclav Paces, Ladislav Pivec & Albert H. Teich (eds.), Science Evaluatiion and Its Management (36-50). Ohmsha: IOS Press, NATO Science Series: Science & Technolgy Policy, vol. 28.

The paper reviews several recent approaches developed for evaluating research programmes in the United States. Specifically, evaluation of three types of programmes will be described: university-based technology transfer offices, exemplified by Iowa State University's Center for Advanced Technology Development; university-based research consortia, exemplified by the Engineering Research Centres (ERC) supported by the National Science Foundation, and the Office of Basic Energy Sciences (OBES) in the U.S. Department of Energy, which supports basic research related to the Department's mission.

Roessner, J. D. (1989). Evaluating Government Innovation Programs: Lessons from the U. S. Experience. Research Policy, 18, 343-359.

Roessner, J. David & Melkers, Julia (1997). Evaluation of National Research and Technology Policy Programs in the United States and Canada. Journal of Evaluation and Program Planning, 20(1).

Roessner, J. David & Melkers, Julia (1995). Evaluation of National Research and Technology Policy Programs in the United States and Canada. In: Stefan Kuhlmann & Doris Holland (Hrsg.), Evaluation von Technologiepolitik in Deutschland: Konzepte, Anwendung, Perspektiven (243-296). Heidelberg: Physica-Verlag.

Roessner, J. David & Melkers, Julia (1992). Evaluation of National Research and Technology Policy Programs in the United States and Canada. Report to the Fraunhofer-Institut für Systemtechnik und Innovationsforschung.

The report describes selected aspects of the evaluation of national science and technology policy programs in the United States and Canada. In the first section of the report, the authors describe the political and structural setting of program evaluation in the U.S. and Canada. In the second and third sections the authors sketch the kinds of evaluations that are done in each country - expecially their methodologies and extent of use - in different types of research and technology programs: fundamental research, medium-term technology development, support of technical and organizational changes, support of industrial innovation, and „experimental" progams. The fourth and fifth sections present case examples of „interesting recent evaluations that span these types of programs. Evaluations were selected because of the economic or political significance of the program itself, the evaluation methods employed, and/or the results (or lack of results) of the evaluation.

Rosenberg, Nathan & Nelson, Richard R. (1996). The Roles of Universities in the Advance of Industrial Technology. In Richard S. Rosenbloom & Spencer, William J. (eds.), Engines of Innovation: U.S. Industrial Research at the End of an Era (87-109). Boston, Massachusetts: Harvard Business School Press.

The last fifteen years have seen growing interest in two questions: How does university research relate to technical advance in industry; and how can American university research become a more effective contributor to the competitiveness of American industry? Is is estimated that about 19 percent of university research now occurs in programs that are linked with industry in some essential matter.

Rosenberg, Nathan & Nelson, Richard R. (1994). American Universities and Technical Advance in Industry. Research Policy, 23, 323-348.

Roussel, Philip A., Saad, Kamal N. & Erickson, Tamara, J. (1991). Third Generation R&D: Managing the Link to Corporate Strategy. Boston, MA: Harvard Business School Press. (ISBN 0-87584-252-6)

The book describes how research and development can be management effectively in a large, complex enterprise to support and enrich its business strategy. The authors characterize this process as third generation management, to distinguish it from the primitive handsoff „strategy of hope" or the somewhat more systematic but incomplete project-management approach. Third generation research and development management is a continuous interactive process. It demands active dialogue and a sense of partnership. In the third generation, guidelines for measuring results and progress are rooted in the principle of management by objectives.

Sellen, Mary K. (1993). Bibliometrics: An Annotated Bibliography, 1970-1990. New York: G. K. Hall & Co.

The purpose of the bibliography is to organize bibliometrical studies to aid those interested in a specific literature (e.g., bibliometric studies in physics) and to aid collection development in libraries. The following observations are made: (1) Science has led and continues to lead the momentum for bibliometric studies. Much of the recent research in scientific bibliometrical studies has been on the research level. These have been conducted to assist in the development of national science policy decisions; determine how productive scientific communities are in a particular country; assess the status of authors and specific academic departments in research universities;

and determine the economic viability of expensive journal titles. (2) The interest in the Social Sciences is growing. As in the Sciences, the majority of studies have been done on the research level. Assessing the status of authors, institutions, and concepts has been a major concern. (3) Interest in the Arts and Humanities is minimal. The few studies in this bibliography indicate no consistent pattern for the studies. (4) Foreign interest is great and primarily limited to scientific literature. Interest in the subject is found in many European and South American publications. Significant studies are found in Japanese journals.

Shadish, William R., Tolliver, Donna, Gray, Maria & Gupta, Sunil K. Sen (1995). Author Judgements about Works They Cite. Social Studies of Science, 25, 477-498.

In general, highly cited scholarly works are rated as exemplars and as being of higher quality. Works rated as highly creative had mixed fates.

Shapira, P., Youtie, J. & Roessner, J. D. (1996). Current Practices in the Evaluation of US Industrial Modernization Programs. Research Policy, 25(2), 185-214.

Small, Henry (1999). Visualizing Science by Citation Mapping. Journal of the American Society for Information Science, 50(9), 799-813.

Spasser, M. A. (1997). Mapping the terrain of pharmacy: co-classification analysis of the International Pharmaceutical Abstracts Database. Scientometrics, 39(1), 77-97.

Stamps, Arthur E. III (Ed.). (1997). Advances in Peer Review Research (Special Issue). Science and Engineering Ethics, 3(1), 1-104.

Peer review is a topic of considerable concern to many researchers, and there is a correspondingly large body of research on the topic. This research raises two basic questions: (a) how does current peer review operate, and (b) how can it be improved? Topics addressed include descriptions of how peer review is used in Federal agencies, whether peer review leads to better manuscripts, demographic characteristics of authors or reviewers, blinding of reviewers, authors, or results, reliability and consistency of reviews, accepting a paper before the study is done, simultaneous submission, and use of dispute resolution procedures such as scientific dialectical and pleading protocols.

Stokes, Donald E. (1997). Pasteur's Quadrant: Basic Science and Technological Innovation. Washington, DC: Brookings Institution Press. (ISBN 0-8157-8177-6)

More than fifty years ago, Vannevar Bush released his enormously influential report „Science, the Endless Frontier“ which asserted a dichotomy between basic and applied science. Stokes challenges Bush's view and maintains that we can only rebuild the relationship between government and the scientific community when we understand what is wrong with that view. Stokes begins with an analysis of the goals of understanding and use in scientific research. He recasts the wideley accepted view of the tension between understanding and use, citing as a model case the fundamental yet use-inspired studies by which Louis Pasteur laid the foundations of microbiology a century ago. During the last decades, technology has been increasingly science based - with the choice of problems and the conduct of research often inspired by societal needs. On this revised, interactive view of science and technology, Stokes builds a convincing case that by recognizing the importance of use-inspired basic research be can frame a new compact between science and government.

Suter, Larry E. (1997). United States: The Experience of the NSF’s Education and Human Resources Directorate. In OECD (Ed.), The Evaluation of Scientific Research: Selected Experiences (107-112). Paris: OECD/GD(97)194.

Teich, Albert H. (1994). Priority-setting and economic payoffs in basic research: An American perspective. Higher Education, 28, 95-107.

Teich, Albert H., Nelson, Stephan D., McEnaney, Celia & Drake, Tina M. (Eds.). (1999). AAAS Science and Technology Policy Yearbook 1999 (Part 4: Evaluating Investments and Performance in Research). Washington, DC: American Association for the Advancement of Science.
http://www.aaas.org/spp/yearbook/

The most important issue in evaluating our national investment in science today is the Government Performance and Results Act (GPRA). This section examines how the GPRA is being implemented. Because it is a new process, growing pains are inevitable and each agency must work to refine its approach. But the GPRA will be with us for the foreseeable future and we must find a way to implement it so we accomplish the true intent of the Act. We also see in this section how the United Kingdom is dealing with assessment and accountability.
Contents
Part 4: Evaluating Investments and Performance in Research
12. Performance of R&D, Accountability of R&D, and the Government Performance and Results Act (Andrew J. Vogelsang)
13. Performance and Accountability: Applying GPRA to Research (Joshua Gotbaum)
14. The Here and Now of NSF and GPRA (Joshua Bordogna)
15. Evaluating Investments and Performance in UK Science (Ben Martin)
16. Results and Responsibility: Science, Society, and GPRA (Susan E. Cozzens)

Vonortas, N. S. (1995). New directions for US science and technology policy: the view from the R&D assessment front. Science and Public Policy, 22(1), 19-28.

Welljams-Dorof, Alfred (1997). Quantitative Citation Data as Indicators in Science Evaluation: A Primer on Their Appropriate Use. In Mark S. Frankel & Jane Cave (Eds.), Evaluating Science and Scientists: An East-West Dialogue on Research Evaluation in Post-Communist Europe (202-211). Budapest: Central European University Press.

The author describes the bibliometric databases of the Institute for Scientific Information (ISI), a primary source for indicators used in the evaluation of research. The value of citation analysis is its use as an indicator of the impact of an author, paper or journal on the scientific community. Dorof outlines some basic rules for the proper use of citation indicators. Using larger citation data increases the confidence level of the results. When ranking units according to performance, the relative trend of performance over time is more significant than a ranking at any given point in time. And establishing a baseline against which performance can be measured and assessed reveals whether performance trends are ahead, equal to or below clearly defined standards. He points out that citation data should always be used in combination with the expert subjective judgments of peers and specialists in the field.

Werner, Bjorn M. & Souder, William E. (1997). Measuring R&D Performance - State of the Art. Research Technology Management, 40(2), 34-42.

Many R&D performance measurement techniques have been developed in response to the unique needs of various organizations. An extensive search of the literature from 1956 to 1995 identified over 90 articles, 12 books and two research reports describing various techniques. Integrated metrics that combine several types of quantitative and qualitative measures were found to be the most effective, but also the most complex and costly to develop and use. The choice of an appropriate R&D measurement metric depends on the user's needs for comprehensiveness of measurement, the type of R&D being measured, the available data, and the amount of effort the user can afford to allocate to it. Guidelines are provided for selecting an appropriate measurement method within these parameters.

Hochschulevaluation USA

Abt Associates (1997). Monitoring of the Graduate Research Traineeship Programme. Cambridge, Massachusetts: Abt Associates, Inc.

Borden, Victor M. H. & Banta, Trudy W. (Eds.). (1994). Using Performance Indicators to Guide Strategic Decision Making. San Francisco: Jossey-Bass, New Directions for Institutional Research No. 82.

The book provides a list of measures that have been cited as examples of performance indicators (pp. 107-117).

Dill, David D. (1998). Evaluating the ‚Evaluative State': implications for research in higher education. European Journal of Education, 33(3), 361-377.

Ferris, James M. (1991). Contracting and Higher Education. Journal of Higher Education, 62(1), 1-23.

The article examines three different applications of contracting to higher education. Despite the various connotations of contracting, they all attempt to increase the efficiency of higher education, either at the system or institutional level. The sources of such gains, however, vary among the different contracting forms. The production cost savings or contracting derive from various factors: scale economies, managerial incentives, managerial flexibility, and incentives for cost-minimization induced from competitive forces. However, such transactions are not costless. Contract administration to minimize potential abuses, that is, adverse selection and moral hazard problems, can be significant.

Franzosa, Susan Douglas (1996). The evaluation of the higher education system in the United States of America. In Robert Cowen (Ed.). World Yearbook of Education 1996: The Evaluation of Higher Education Systems (126-143). London: Kogan Page.

Glassick, Charles E., Huber, Mary Taylor & Maeroff, Gene I. (1997). Scholarship Assessed: Evaluation of the Professoriate. San Francisco: Jossey-Bass Publishers.

The book examines the changing nature of scholarship in today's colleges and universities. It proposes new standards for assessing scholarship and evaluating faculty with special emphasis on methods for documenting effective scholarship.

Goldberger, Marvin L., Maher, Brendan A. & Ebert Flattau, Pamela (Eds.). (1995). Research-Doctorante Programs in the United States – Continuity and Change. Washington, DC: National Academy Press.

Kells, H. R. (1999). National higher education evaluation systems: Methods for analysis and some propositions for the research and policy void. Higher Education, 38, 209-232.

A review of the status of national higher education evaluation systems and an examination of the lack of research into the process and nature of policy choice which brought them into existence and shaped them. Proposals are made concerning methods to assist future policy choices and analysis of these systems and propositions posited to further our understanding and as a basis for further research.

Kells, H. R. (1996). Self-Study Processes: A Guide to Self-Evaluation in Higher Education (4th edition). New York: American Council on Education / Oryx Press.

An independent consultant on higher education explains how to derive maximum benefit from the self-analysis process to improve programs. The focus is on American institutions but the author provides multinational perspective, as well.

El-Khawas, Elaine (1998). Strong State Action but Limited Results: Perspectives on University Resistance. European Journal of Education, 33(3), 317-330.

Mansfield, Edwin (1995). Academic research underlying industrial innovations: sources, characteristics, and financing. Review of Economics and Statistics, 57, 55-65.

There has been no systematic study of the characteristics of the universities and academic researchers that seem to have contributed most to industrial innovation. Nor do we know how such academic research has been funded. The paper, based on data obtained from 66 firms in seven major manufacturing industries and from over 200 academic researchers, sheds new light on the sources, characteristics and financing of academic research underlying industrial innovation.

Mansfield, Edwin & Lee, J.-Y. (1996). The modern university: contributor to industrial innovation and recipient of industrial R&D support. Research Policy, 25, 1047-1058.

Massy, William F. (Ed.). (1996). Resource Allocation in Higher Education. Ann Arbor, MI: The University of Michigan Press.

This book lays the groundwork needed for institutional leaders and government officials to assess their existing resource allocation processes and plan for improvement. The book presents examples of decentralized resource allocation processes and process elements from higher education. A budgeting approach is proposed that combines the positive aspects of Performance Responsibility Budgeting (PRB) and Revenue Responsibility Budgeting (RRB) while skirting their main problems. The book examines also the linkage between resource allocation reform and academic restructuring at the grassroots level, a connection that must be planned carefully lest departments fall into the quality trap.

Mets, L. A. (1995). Programme review in academic departments. New Directions for Institutional Research, 86, 19-36.

Nettles, Michael T., Cole, John J. K. & Sharp, Sally (1997). Benchmarking Assessment - Assessment of Teaching and Learning in Higher Education and Public Accountability: State Governing, Coordinating Board & Regional Accreditation Association Policies and Practices. Stanford, CA: National Center for Postsecondary Improvement, Stanford University, School of Education.
http://ncpi.stanford.edu

The report presents the results of the first of four stages of research to be conducted from 1996 through 2001. The research project aims to explore the progress that has been made by the 50 states and six regional accrediting associations toward establishing and implementing higher education assessment policies. The primary interest is in policies and practices that seek to improve teaching and learning in the nation's colleges and universities. The rationale offered by accreditation associations and the states for adopting assessment practices has varied, but there are some common themes/phrases that emerge across the nation, including the following:
• increasing public accountability to taxpayers whose taxes provide the largest single source of funding for colleges and universities;
• ensuring quality to citizens by providing concrete evidence about the instructional performance of the colleges and universities that they are considering attending or otherwise supporting;
• identifying strengths and limitations of colleges and universities for purposes of state planning;
• achieving greater efficiencies in state systems of higher education and within individual institutions;
• identififying new criteria to use in funding colleges and universities; and
• increasing international, interstate, and intra-state competition for high quality higher education.

Nettles, Michael T. & Cole, John J. K. (1999). State Higher Education Assessment Policy: Research Findings from Second and Third Years. Stanford, CA: National Center for

Postsecondary Improvement, Stanford University, School of Education.
http://ncpi.stanford.edu

A questionnaire (The State Higher Education Assessment Questionnaire) was mailed to all 50 state academic officers to explore the dynamics of the policy process by which assessment becomes a state-level issue or concern, and to understand better the mulifaceted relationship between state assessment policy on one hand and the improvement of teaching and learning on the other.

Rudasill, Lynne (1997 ff.). College and Universities Rankings. Urbana, IL: Education and Social Science Library University of Illinois at Urbana-Champaign.
The URL for the U. S. ranking web:
http://www.library.uiuc.edu/edx/rankings.htm
College rankings bibliography:
http://www.library.uiuc.edu/edx/rankbib.htm

From sports to education, Americans are captivated by rankings. Everyone wants their favorite team to be the best, and every parent wants their child to attend the best college or university. While sports teams can compete on the playing field to determine who is better on a given day, institutions of higher education have no commonly agreed upon measures to allow for the comparison of teaching and research programs. The overwhelming variety in size and purpose of institutions makes the prospect of comparing colleges a daunting one for students, parents, and the colleges themselves.
For many years, various bodies have undertaken statistical and reputational rankings of colleges and attempted to provide information to prospective students. Increasingly, the importance and validity of college rankings is a hotly debated issue. Many universities, including highly ranked ones, are beginning to question both the data and methods used by some ranking services. Of special concern are the aspects of the rankings which deal with the difficult to measure concept of institutional reputation. At this site, you will find links to many ranking services, along with cautionary notes and a discussion of the ongoing controversy over rankings.
The listing of items on our site's pages in no way constitutes an endorsement of a ranking service by the Education and Social Science Library or by the University of Illinois at Urbana-Champaign. The purpose of our site is only to draw together and provide context to various ranking services. Hopefully, the information found here will improve your knowledge about rankings in general and help you better use these services. Toward that end, we highly encourage you to peruse our Caution and Controversy page. Additionally, our College Rankings Bibliography provides numerous articles on the topic.

Trow, Martin (1998). On the Accountability of Higher Education in the United States. In: William G. Bowen & Harold T. Shapiro (Eds.), Universities and Their Leadership (15-61). Princeton, NJ: Princeton University Press.

Trow, Martin (1994). Academic Reviews and the Culture of Excellence. Stockholm: Kanslerämbetets skriftserie (1994: no. 1).

The author distinguishes four kinds of review: (1) internal supportive, (2) internal evaluative, (3) external supportive, and (4) external evaluative. He stresses the value of internal supportive reviews for the maintenance of academic quality.

Walleri, R. Dan & Moss, Marsha K. (Eds.). (1995). Evaluating and Responding to College Guidebooks and Rankings. San Francisco: Jossey-Bass Publishers (New Directions for Institutional Research, Number 88).

The volume explores some of the major facets of and issues surrounding college guidebooks and ratings. The background and development of these publications are traced, followed by discussion of major issues and perspectives - consumer use of the publications, validity of ratings, and the institutional burden of supplying the needed information. Views from both the institutions and the publishers are presented.

Webster, David S. & Skinner, Tad (1996). Rating PhD Programs: What the NRC Report Says and Doesn't Say. Change, 28(3), 34-44.

Zamarripa, Edward J. (1995). Evaluating Research Productivity. SRA - Journal of the Society of Research Administrators, 1995, 26 (3-4), 17-27.

This study compares the attitudes of research scientists with those of research administrators to (1) determine whether the two groups differ in their perceptions of research productivity, and (2) suggest which measures might be used in assessing research productivity. The results include a rank-order comparison of the relative importance of several items that measure research productivity as perceived by scientists and administrators.

3. Dänemark

Forschungsevaluation

Christensen, F. Hjortgaard, Ingwersen, P. & Wormell, Irene (1997). Online determination of the Journal Impact Factor and its international properties. Scientometrics, 40(3), 529-540.

Christensen, F. Hjortgaard & Ingwersen, P. (1996). Online citation analysis: A methodological approach. Scientometrics, 37(1), 39-62.

Danish Council for Research Policy (1993). International Evaluation of Danish Health Research. Copenhagen: Danish Council for Research Policy.

Danish Council for Research Policy (1992). International Evaluation of Danish Agricultural Research. Copenhagen: Danish Council for Research Policy.

Danish Research Councils (1998). The Materials Technology Development Programme - Evaluation of Research and Technology Relevance. Copenhagen: The Danish Research Councils.
http://www.forskraad.dk/snf/publikation/mup2-indhold.htm
http://www.forskraad.dk/publ-uk.html

The present evaluation of MUP-2 consisted of two steps. Firstly, the different sub-programmes (Centres, Framework Programmes and Demonstration Projects) prepared self-evaluation reports that were reviewed by a large group of international scientific experts. Secondly, an International Scientific Panel (ISP) comprised of seven of these experts conducted site visits to specific Centres and Framework Programmes selected by the Coordination Committee. The members of the ISP representing different research interests within the field of materials science and technology were selected so as to cover the range of MUP-2 research activities and ensure the necessary industrial experience. The members of the ISP are listed in Annex 1 and the site visit schedule programme is given in Annex 2. The present Evaluation Report summarizes the ISP's impression of MUP-2 gained from (1) the self-evaluation reports, (2) the comments to these self-evaluation reports and (3) the site visits.

Danish Research Councils (1997). Evaluation of FØTEK. The Steering Committee's Summary Report. Copenhagen: The Danish Research Councils.
http://www.forskraad.dk/publ/fotek/

In 1996, it was decided to initiate a broad evaluation of The Danish Research and Development Programme for Food Science and Technology (FØTEK). In the course of the evaluation, four sub-

reports were published which collectively described the total research, educational and innovative results and provided recommendations for use in connection with decisions on future activity. In the present report the Steering Committee summarises the results and presents its recommendations to the participating ministries.

Danish Research Councils (1995). Midterm Evaluation of Centres of Engineering Science and framework programmes. Copenhagen: The Danish Research Councils.

Danish Research Councils (1993). Mid-term Evaluation: Danish Biotechnological Research and Development Programme (1991-1995). Copenhagen: The Danish Research Councils.

Danish Research Councils (1993). Mid-term Evaluation: Danish Research Programme on Informatics. Copenhagen: The Danish Research Councils.

Ingwersen, Peter (1999). On-line indicators of Danish biomedical publication behaviour 1988-96: international visibility, impact and co-operation in a Scandinavian and world context. Research Evaluation, 8(1), 39-45.

The aim of the paper is to demonstrate seven central publication indicators and determine changes in research behaviour and policy over time in a domain. Scandinavia's position is under pressure and weakening, in publication and citation impact levels. Only Finland demonstrates a steady increase in research activity. The Danish position is stable but the pattern of research publication is changing dramatically because of a continuous shift to publish in Science Citation Index journals and an intensive enhancement of international collaboration, in range of co-operating countries and in volume.

Ingwersen, Peter & Christensen, Finn H. (1997). Data set isolation for bibliometric on-line analysis of research publications: fundamental methodological issues. Journal of the American Society for Information Science, 48(3), 205-217.

Stahle, Bertel (Ed.). (1987). Evaluation of Research: Nordic Experiences. Copenhagen: Nordic Science Policy Council, FRP-publication no. 5.

Hochschulevaluation Dänemark

Analyseinstitut for Forskning (Ed.). (1999). Assessing Assessments - European Experiences. Proceedings of a conference organized by the Danish Institute for Studies in Research and Research Policy in cooperation with The European Consortium for Political Research. Aarhus: The Danish Institute for Studies in Research and Research Policy.
http://www.afsk.au.dk/ftp/assessment/assessment.pdf
http://www.afsk.au.dk

The seminar was planned within a framework determined by the review process, discussing questions such as: who commissions the review (formal, Ministry or informal), who appoints the review panels and determines the brief, who sits on the review panel, how do they do their work (visits or paperwork alone), to whom they report (public or private), what are the implications and whether there is a feed-back effect? In the report European experiences within research assessment is described and discussed. The focus is primarily at political science but the issue whether experiences from political science assessments could be transferred to other fields of science was on the agenda as well at the seminar. To give an understanding of the national differences in organising assessments and evaluations the practice is described for more than eight European countries including the Netherlands, Belgium, Italy, Spain, Germany and Greece, not to forget the British and the Irish experiences which has highly influences the rest of Europe. Finally the issue whether the results of assessments can be used in practice was discussed at the seminar.

Foss Hansen, Hanne & Borum, Finn (1999). The Construction and Standardization of Evaluation. The Case of the Danish University Sector. Evaluation, 5(3), 303-329.

The article is based on an empirical study of research evaluation and evaluation of education within the Danish university sector. It conceptualizes and explains similarities and differences in adoption processes and constructed evaluation standards within two subfields. Three models are used for explaining differences and similarities: a stakeholder model, an institutional field model and an institutional heritage model. The article shows how evaluation as an organizational element is used simultaneously for processes of change and processes of reproduction of norms and values.

Foss Hansen, Hanne (1995). Organizing for Quality - A Discussion of Different Evaluation Methods as Means for Improving Quality in Research. Science Studies, 8(1), 36-43.

Thune, Christian (1999). Denmark Launches a Single Organization for the Evaluation of All Levels of Education.
http://www.evc.dk/publika/publ_artik.html

In May 1999 the Danish parliament passed a law proposed by the government and providing the legal background for a new institution, The Danish Institute of Evaluation. The mandate of the Institute is internationally unique, because it is given the task by parliament to undertake systematic and mandatory evaluation of teaching and learning at all levels of the educational system from kindergarten classes to post graduate programmes.

Thune, Christian (1997). The Balance Between Accountability and Improvement: The Danish Experience. In John Brennan, Peter de Vries & Ruth Williams (Eds.), Standards and Quality in Higher Education (87-103). London: Jessica Kingsley Publishers.

The Danish Centre for Quality Assurance and Evaluation of Higher Education (http://www.evc.dk/)was established by the Ministry of Education in 1992. The Centre is in principle independent of the Ministry of Education, and of the universities and other institutions of higher education. The mandate of the Centre is: to initiate evaluation processes of higher education in Denmark; to develop appropriate methods of assessing programmes; to inspire and guide the institutions of higher education in aspects concerning evaluation and quality; and to compile national and international experience on evaluation of the educational system and quality development. A substantial part of the Centre's work consists of regular and systematic evaluations of programmes on a rotating basis in which all programmes will be evaluated within a period of seven years (List of evaluation reports: http://www.evc.dk/publika/publ_eval.html). In addition, the Centre evaluates new programmes after their establishment period, and programmes for which the Ministry of Education, consulting bodies, or an HEI find that there is a need for an evaluation of the quality of the programme. The Centre must ensure that reliable methods are employed in connection with the execution of the various evaluations.

4. Grossbritannien

Forschungsevaluation

Advisory Board for the Research Councils (1990). Peer Review. London.

The Working Group on Peer Review concluded that there was no practicable alternative to peer review for the assessment of basic research and that present systems for its operation were essentially satisfactory. But these systems were seen to be under strain and in need of review to improve their effectiveness and to sustain the confidence of the academic community. The working group's report set out detailed recommendations for achieving these objectives.

Arnold, Erik & Balázs, Katalin (1998). Methods in The Evaluation of Publicly Funded Basic Research. A Review for OECD. Brighton: Technopolis Ltd.
A copy of this report is available from:
http://www.technopolis.co.uk/reports

Barber, John M. (1999). Creating an Anglo-Saxon Innovation Culture. In: Susanne Bührer & Stefan Kuhlmann (Eds.), Evaluation of Science and Technology in the new Europe. Proceedings of an International Conference on 7 and 8 June 1999, Berlin (33-44). Bonn / Bruxelles: Federal Ministry of Education and Research / European Commission.

Barker, Derek & Lloyd, Philippa (1997). Evaluation of Scientific Research in the United Kingdom. In OECD (Ed.), The Evaluation of Scientific Research: Selected Experiences (47-58). Paris: OECD/GD(97)194.

Boddington, Andy (1993). Research evaluation systems: sources of policy formation and 'evaluation push'. Research Evaluation, 3(3), 197-203.

Research evaluation is not only conducted through ad hoc projects but also through large-scale, long-term, systematic evaluations of entire sectors or activities. These are denoted 'research evaluation systems'. One characteristic of such systems that has not been studied sufficiently is 'evaluation push' - the process through which evaluation criteria influence the conduct and performance of research. Two UK research evaluation systems, operated by the ESRC (Economic and Social Research Council) and the former UFC (Universities Funding Council), have distinct evaluation push characteristics. Both systems, and the UFC one in particular, are also major sources of information for policy-makers.

Brook, Richard (1997). Policy Making and Peer Review in the UK Engineering & Physical Sciences Research Council. In: Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Ed.), The Future of the Peer Review System (7-25). The Hague: NWO.

Within the EPSRC, new methods for the setting of research priorities and new methods for the selection of particular research projects matched to those priorities were introduced in 1994. A description is given of the operation of the new system; it indicates something of the reasons for the changes that have been made, it describes the mechanisms that are involved in the new procedures and it summarises initial experiences with the system.

Cunion, K. M. (1995). U. K. Government departments experience of RT&D programme evaluation and methodology. Scientometrics, 34(3), 375-389.

The UK Department of the Environment is responsible for a range of policy issues within Government related to many aspects of the environment in its broadest sense. Over recent years a system of research assessment has been established which consists of the development of ROAME statements (cf. http://www.miti.go.jp/info-e/cM99809e.html) for the appraisal of programmes and

regular independent evaluation of the success and impact of the research on the basis of a five year cycle (ROAME = **R**ationale, **O**bjectives, **A**ppraisal, **M**onitoring, **E**valuation, Feedback). The mechanisms and process of the assessment system are described. Effective evaluation of policy-oriented research programmes has provided valuable information to the Department on the success and impact of research, and guidande on future direction and balance of the programme.

Cunningham, P. N., Georghiou, L. G. & Barker, K. E. (1992). Re-orienting evaluation in a research council. Research Evaluation, 2(2), 111-118.

The Social Sciences and Humanities Research Council of Canada (SSHRC) proposed to re-orient its evaluation function to move away from a programme-based, objective-oriented system towards a more strategic approach looking at the wider and longer-term impacts. This study by PREST recommended a combination of the development of facilitating tools for evaluation which can be used at several levels, and strategic studies.

David, P. A. (1997). From market magic to calypso science policy - A review of Terence Kealey's „The economic laws of scientific research“. Research Policy, 26, 229-255.

Department of Trade and Industry (o. J.). The ROAME System of the UK Department of Trade and Industry (DTI). London.
http://www.miti.go.jp/info-e/cM99809e.html

Economic and Social Research Council (1998). Evaluation Guide One: Evaluating End of Award Reports. London: ESRC, Policy and Evaluation Division.

Economic and Social Research Council (1998). Evaluation Guide Two: Evaluating Research Programmes. London: ESRC, Policy and Evaluation Division.

Economic and Social Research Council (1997). Evaluation Guide Third: Evaluating Research Centres. London: ESRC, Policy and Evaluation Division.

Freeman, Chris & Soethe, Luc (1997). The Economics of Industrial Innovation (Chapter 10: Uncertainty, Project Evaluation and Innovation, pp. 242-263). London: Pinter, Third Edition.

Georghiou, Luke (1999). Meta-Evaluation: Evaluation of evaluations. In: European Commission & Austrian Advisory Board for Universities (Eds), Science and the Academic System in Transition: An International Expert Meeting on Evaluation (195-202). Budapest: Akadémiai Kiadó.

Criteria to assess evaluations: Studies of user needs from evaluations performed by PREST have shown that there are three major dimensions to be considered. The most important is that of the credibility of the evaluation. A second dimension of user satisfaction is that of the absorbability of the evaluation - that of how the report is presented. A third dimension is outside the evaluation, concerning rather how it interfaces with the decision-making process.

Georghiou, Luke (1998). Issues in the evaluation of innovation and technology policy. Evaluation, 4(1), 37-51.

Georghiou, Luke (1995). Research evaluation in European national science and technology systems. Research Evaluation, 5(1), 3-10.

The practice of research evaluation in the European Union differ as would be expected given the diversity of systems in which evaluation is carried out. While programme evaluations are becoming more routine, institutional reforms have created a demand for a new kind of evaluation. This includes a strategic orientation which is likely to become a regular feature of more important programme evaluations. Trends towards infrastructural innovation policies will inevitably lead to an increased demand for evaluations of initiatives in this sphere.

Georghiou, Luke & Roessner, David (1999). Evaluating Technology Programs: Tools and Methods. Manchester: University of Manchester, PREST, and Atlanta, GA: Georgia Institute of Technology. Unpublished Manuscript.

The first focal point of the paper concerns evaluation of publicly-supported research carried out in universities and public-sector research organizations. One reason this is included is the growing significance attached to the economic and social rationales for public support of research. Hence, the paper addresses means by which the economic and social value of science is being assessed. In the second part of the paper the scope is broadened to include evaluations that focus upon linkages, including those of programs seeking to promote academic-industrial and public-private partnerships. Finally, experience in the evaluation of diffusion and extension programs is discussed.

Gibbons, Michael (1985). Methods for Evaluation of Research. International Journal of Institutional Management in Higher Education, 9, 79-85.

Gibbons, Michael & Georghiou, Luke (1987). Evaluation of Research - A Selection of Current Practices. Paris: OECD.

Guy, Ken (1998). Strategic Options for the Evaluation of the R&D Programmes of the European Union. Final Report. Prepared for STOA. Brighton: Technopolis Ltd.
A copy of this report is available from:
http://www.technopolis.co.uk/reports

Guy, Ken & Arnold, Erik (1995). UK Government practice in science and technology evaluation. In: Stefan Kuhlmann & Doris Holland (Hrsg.), Evaluation von Technologiepolitik in Deutschland: Konzepte, Anwendung, Perspektiven (297-316). Heidelberg: Physica-Verlag.

There has been considerable development in the methods and use of evaluation in British science, technology and industry policy since 1980. Evaluation is beginning to move from testing whether programmes perform against their allotted tasks to more strategic questions about programme appropriateness. This implies increased selectivity in the use of certein evaluation resources in order to focus on policy-relevant issues. At the same time, evaluation's critical roles in organisational learning and as a key part of the management information system of agencies continue to be recognised.

Guy, Ken & Arnold, Erik (1993). UK Government practice in science and technology evaluation. Research Evaluation, 3(3), 179-186.

In the UK the purpose of and approach to evaluation of state activity in science and technology has changed over recent years from a tactical tool, to determine whether programmes have performed their allotted tasks, to a strategic one which investigates the programmes appropriateness. By focusing on three agencies (DTI, SERC, ESRC), the development in the methods and use of evaluation is traced and the importance of evaluation for policy-making emphasised.

Guy, Ken, Georghiou, L., Quintas, P., Cameron, H., Hobday, M. & Ray, T. (1991). Evaluation of the Alvey Programme for Advanced Information Technology. London: HMSO.

Halfpenny, Peter & Miles, Ian (1993). Evaluating interdisciplinary social science initiatives: experiences from the UK. Research Evaluation, 3(3), 134-150.

The problems involved in evaluating interdisciplinary social science initiatives are described, based on experience appraising large ESRC-funded projects undertaken in the UK. The methods used include interview and questionnaire approaches addressed to investigators, the research community, and potential users of the research. While there was substantial consensus among

different groups as to the achievements of the Initiative, especially in terms of the data produced, there was greater diversity in the assessment of the intellectual contribution and policy use of the results. This reflected the diversity of disciplinary, methodological and institutional locations of the assessors. Such diversity is more often than not going to be the rule when interdisciplinary research is involved. This has significant implications for evaluation practice.

Hicks, Diana & Katz, Sylvan (1997). The Changing Shape of British Industrial Research. Falmer, Brighton: University of Sussex, SPRU (STEEP Special Report No 6). http://www.sussex.ac.uk/spru/news/pressrel/besstsum.html

The Bibliometric Evaluation of Sectoral Scientific Trajectories (BESST): Phase II project analysed 41,000 scientific and technical articles published by UK industry. The data analysed encompass:
- fourteen years - 1981 to 1994
- all UK companies publishing in journals
- indexed in the Science Citation Index (SCI)
- all science fields, but not management, social sciences or humanities
- all collaborations by companies
- citations
research output of a certain quality, commonly thought of as academic research.

Hicks, Diana (1991). Hospitals: the hidden research system. Science and Public Policy, 23(5), 297-304.

Hicks, Diana (1991). A cautionary view of co-citation analysis. Research Evaluation, 1(1), 31-36.

Co-citation analysis requires further development and problems remain with current maps. With the help of seven basic questions it is hoped to allow a more critical examination of these maps and to bridge the gap between the true believers and the radical critics.

Hills, Philip V. (1999). Research Evaluation and Management in the UK Government. In Vaclav Paces, Ladislav Pivec & Albert H. Teich (eds.), Science Evaluatiion and Its Management (127-137). Ohmsha: IOS Press, NATO Science Series: Science & Technolgy Policy, vol. 28.

The paper describes the UK Government's involvement in research and its approach to managing and evaluating that involvement. Evaluation in UK government departments is heavily impact oriented and to a fairly large extent programme based. It is carried out with a fair degree of independence and objectivity. Methodology varies mainly in relation to the nature of the research involved. Where research is relatively basic, scientific excellence is the main criterion and in all cases it is a vital factor. In the case of research of a more applied kind, with shorter term economic effects, there is emphasis on indicators of that impact, and evaluation draws heavily on the views of users and/or customers. The procedures adopted range, therefore, from traditional peer review to heavily survey oriented investigation.

Hills, Philip V. (1995). PREST's experience of evaluation. Scientometrics, 34(3), 401-414.

Hills, Philip V. & Dale, Alison J. (1995). Research and technology evaluation in the United Kingdom. Research Evaluation, 5(1), 35-44.

It can be argued that the UK has one of the most advanced research evaluation systems in Europe. This system has developed from a means of ensuring value for money in public expenditure, to a process aimed at informing decision-making at all levels, including that of policy-making. This paper describes the UK's evaluation system in the context of the country's science and technology policy, in terms of the actors, institutions and methodologies involved. It discusses the most recent developments in this field, and asks whether or not they address the problems that still beset the evaluation process, even after many years of practice.

Katz, J. Sylvan & Plevin, John (1998). Environmental science in the UK: a bibliometric study. Research Evaluation, 7(1), 39-52.

Scientific publications in the peer-reviewed open literature are a major output of research and a key mechanism in the dissemination of knowledge. The report examines the publication record of selected subject categories with strong links with the environmental sciences in the Institute for Scientific Informatin database. UK publications and citations are compared with the world output and that from a number of overseas countries. A standard set of measures is used to look at the number of papers produced, their impact and the extent of collaboration between UK scientists and colleagues in industry and overseas. UK environmental science emerges as a mature research system with strength across many disciplines and impact levels above the world average. However, for a number of the fields examined, the priority given by the UK to the environmental sciences within its overall science programme appears to be lower than that given by other countries.

Katz, J. Sylvan & Hicks, Diana (1997). Desktop Scientometrics. Scientometrics, 38(1), 141-153.

The paper examines how the BESST project developed a Desktop Scientometric environment using public domain, hardware independent software, prototyped a graphical user interface to provide easy access to UK sectoral level bibliometric data and gives a glimpse into future developments.

Katz, J. Sylvan & Hicks, Diana (1997). Bibliometric Indicators for National Systems of Innovation (IDEA project funded by TSER program of the EC). http://www.sussex.ac.uk/spru/best/nsi/index.html

Classical bibliometrics focuses on the national level and international comparisons. Even with the emerging emphasis on disaggregation, international comparison and analysis of interdependencies will be required, and we illustrate the ease with which national systems can be set in an international context bibliometrically. The sectoral and intra-sectoral level data we have developed are possible due to recent advances in desktop computing. These data can make their most powerful contribution in the context of the new approaches to innovation - although we do not make those connections here (for more detailed efforts in this direction see Hicks and Katz, 1997). For each level, we propose four general types of indicators:
1. size or number of papers, the classical measure of research output;
2. impact or number of citations, again a classical bibliometric indicator; 3. diversity in capabilities derived from size, impact, size growth and impact growth distributions across scientific fields; 4. interaction in research networks as evidenced by collaborative research output and derived using size, impact and diversity measures of co-authored papers.

Kealey, Terence (1996). The Economic Laws of Scientific Research. London: Macmillan Press.

Does academic science breed new technology, and does new technology breed economic growth? Is government funding required to optimise the step? Would the free market supply enough basic science? To help answer these questions, the author surveys the science policies and economic outcomes of two science nations: the USA and the UK. Each was the lead country economically, becoming so while pursuing laissez faire policies for science. Each, however, is now scientifically dirigiste. The book surveys the evolution of their science policies, to explain why they shifted from *laissez faire* to *dirigiste*, and it examines the consequences of the shift. He tries to determine why almost every other major industrialised country is now scientifically *dirigiste*.

Lewison, Grant, Cottrell, Robert & Dixon, Diane (1999). Bibliometric indicators to assist the peer review process in grant decisions. Research Evaluation, 8(1), 47-52.

The Wellcome Trust has been using bibliometrics for the last three years to inform the panel that makes decisions on longer-term research grants in neurosciences. These compare an applicant's

publications with those of a handful of scientific peers, and citations to these papers compared with a norm group in the applicant's subfield. The paper reports three surveys, two of panel members and one of applicants, to determine their knowledge and views of bibliometrics and of which indicators were the most useful. More than two-thirds of the respondents were in favour of using bibliometrics. They considered citations scores and journal-impact category rankings as being the most helpful. The panel has now decided to continue using bibliometric indicators but to simplify the analysis to make it more cost-effective.

Lewison, Grant & Dawson, G. (1998). The effect of funding on the outputs of biomedical research. Scientometrics, 41(1-2), 17-27.

Lewison, Grant (1995). Evaluation of national bibliometric research outputs through journal-based esteem measures. Research Evaluation, 5(3), 225-235.

The study was carried out as part of the UK Government's Technology Foresight Programme to provide an assessment of the current relative strengths and weaknesses of the UK and some 11 other OECD countries in 18 subfields of medical and biomedical research, as part of the panel's work in 'benchmarking' the status of UK biomedicine. The paper is concerned with the techniques used and the numerical results obtained.

MacLean, Marlie, Davies, Catherine, Lewison, Grant & Anderson, Loe (1998). Evaluating the research activity and impact of funding agencies. Research Evaluation, 7(1), 7-16.

While the utility of assessing past research is widely recognised, few studies have focused on the funding body as the unit of evaluation. The study focused on individual funding bodies in one field of research, malaria: the results are presented of a survey of both the international financial inputs and the publications that resulted. Some major funding organisations obtained more apparent productivity from their investment than did others, although the leading funding bodies all supported papers that were more highly cited than the average for the field. The mean number of funding bodies acknowledged on more highly cited papers was greater than that for the complete set of papers, suggesting that the presence of multiple funding is positively correlated with citation performance. Other subjective methods of assessment involving surveys of expert opinion are also discussed.

Martin, Ben R. (1999). Evaluating Investments and Performance in UK Science. In Teich, Albert H., Nelson, Stephan D., McEnaney, Celia & Drake, Tina M. (Eds), AAAS Science and Technology Policy Yearbook 1999 (Part 4: Evaluating Investments and Performance in Research). Washington, DC: American Association for the Advancement of Science. http://www.aaas.org/spp/yearbook/

The author discusses the key features of the UK research system as well as recent science policy in the UK. He also discusses experiences with evaluation in three sets of agencies: the Higher Education Founding Council (and the university research assessment exercises), the research councils, and government department research. He details how the assessment exercises have worked in academia and in Institutes.

Martin, Ben R. (1997). Factors affecting the acceptance of evaluation results. In Mark S. Frankel & Jane Cave (Eds.), Evaluating Science and Scientists: An East-West Dialogue on Research Evaluation in Post-Communist Europe (28-45). Budapest: Central European University Press.

The author presents a number of conclusions about factors that govern the acceptance of evaluation results. The essential first step in an evaluation must be to map out the wider political context, indentifying the key actors and their respective interests. A second conclusion is the importance of involving as fully as possible in the evaluation all those being assessed, be they scientists, policy makers or any other stakeholders. Third, the evaluation methodology must ultimately be acceptable to those being assessed. A fourth conclusion is that the results must be simply and succinctly

expressed. Next, for any evaluation, there are several different audiences who may be interested in the results. For different audiences, one needs different mechanisms for ensuring that the evaluation findings are disseminated as effectively as possible. Finally, the institutional location and degree of autonomy of the evaluators is crucial. The evaluators need to be close - but not too close - to the policy makers if their efforts are to be effective and their results accepted.

Martin, Ben R. (1996). The use of multiple indicators in the assessment of basic research. Scientometrics, 36(3), 343-362.

The paper argues that evaluations of basic research are best carried out using a range of indicators. After setting out the reasons why assessments of government-funded basic research are increasingly needed, we examine the multi-dimensional nature of basic research. This is followed by a conceptual analysis of what the different indicators of basic research actually measure. Having discussed the limitations of various indicators, we describe the method of converging partial indicators used in several SPRU evaluations. The paper also reports the results of a survey of academic researchers. They, too, are strongly in favour of research evaluations being based on multiple indicators combined with peer review. The paper ends with a discussion as to why multiple indicators are not used more frequently.

Martin, Ben R. (1996). Technology foresight: capturing the benefits from science-related technologies. Research Evaluation, 6(2), 158-168.

Technology foresight is a process for bringing together scientists, industrialists, government officials and others to identify the areas of strategic research and the emerging technologies likely to yield the greatest ecoomic and social benefits. The experiences with technology foresight in six countries are summarised. A conceptual model of the forsight process is proposed, and some of the factors structuring success and failure in foresight are analysed.

May, Robert M. (1998). The scientific investments of nations. Science, 281 (3 July), 49-51.

In this paper the Chief Scientific Adviser to the UK Government compares 12 countries' national investment in R&D between 1981 and 1995, using OECD data. To investigate the relative effectiveness of different countries' investment in R&D, the author compares countries' relative investment in basic research with their relative output of scientific research papers.

May, Robert M. (1997). The scientific wealth of nations. Science, 275 (7 February), 793-796.

In this paper the author offers comparisons, from a variety of viewpoints, of scientific research outputs among several countries.

Office of Science and Technology (1997). The Quality of the UK Science Base. London: Department of Trade and Industry, Office of Science and Technology.

Power, Michael (1997). The Audit Society: Rituals of Verification (Auditing Academic Research, pp. 99-101). Oxford: Oxford University Press.

Power, Michael (Ed.). (1996). Accounting and science: Natural inquiry and commercial reason. Cambridge: Cambridge University Press (Cambridge Studies in Management 26).

In recent years policymakers and scientists have become increasingly interested in the economics of science, and in particular in the relationship between accounting and science. The book explores the intersections between the sociology and history of science and the sociology of accounting. The contributors explore a number of issues, including the role of accounting as a distinctive form of administrative objectivity; conceptual exchanges between science and business administration; actuarial practices and their claims to scientificity; conceptions of the factory as a form of laboratory; accounting for research and development expenditures; the emerging role of patents in the physical sciences; and models of scientific accountability.

Royal Society (1996). Peer Review. An Assessment of Recent Developments. London: Royal Society.
http://www.royalsoc.ac.uk/st_pol05.htm

The conduct of peer review in the UK has been the subject of extensive debate since the White Paper „Realising our potential“ was published in May 1993. The debate has been triggered by, inter alia, the major expansion of the university system, the introduction of new Research Councils with new missions in 1994, the explicit focus on wealth creation and quality of life, the introduction of schemes such as ROPAs that use non-traditional assessment systems, and the renewed emphasis on efficiency in the conduct of Research Council business. The Council of the Royal Society therefore appointed a group to consider the changes that have actually taken place and to address concerns about how peer review is operating in the rapidly changing environment of public funding for the Science Base. The Group's remit focused on peer review in the context of judging proposals for funding of projects, programmes or Research Council Units or for fellowship support. It did not extend to the Society's own assessment procedures (which are under separate consideration), nor to decision-making processes in industrial R&D. Peer review (refereeing) of papers submitted for publication was excluded from the remit.

Sherman, Brad (1994). Governing Science: Patents and Public Sector Research. Science in Context, 7(3), 515-537.

While recognizing that public sector research has long been managed by a wide variety of practices and techniques, this paper concentrates on the increasingly important role that patents are playing in the management and regulation of public sector research.

Solesbury, William (1996). Scientific research: demystifying peer review. Research Evaluation, 6(1), 19-23.

In practice peer review focuses on three concerns. 'Fitness for purpose' - if unfit, the science is invalid, certainly not worh funding. 'Knowledge added' - all science worth funding should offer some gain. 'Value for money', where the value is assessed as knowledge added, is the appropriate criterion for making choices with limited budgets. This three-stage decision algorithm can demystify and defend peer-review practice.

Stern, Elliot (1993). Ongoing and participative evaluation: purpose, design and role in the evaluation of a large-scale R&D programme. Research Evaluation, 3(2), 75-82.

Using the particular case of a large-scale R&D programme concerned with learning technologies within the European Community's Third Framework Programme for pre-competitive industrial R&D, a participative evaluation method is outlined. It is important to define adequately the object of evaluation at micro, meso and macro levels; the programme architecture defines who are the legitimate actors and how they are organised and funded; by defining the stakeholders a joint enterprise evaluation was agreed by which the questions of 'local' evaluators were given as much precedence as those of the programme's sponsors; an understanding of the context of the R&D requires a knowledge of the domain.

Tiler, Christine & Boddington, Andy (1993). Outputs, structure and process in the evaluation of social science research centres. Research Evaluation, 3(2), 107-116.

Using the evaluation of two social science research centres as an example, this illustrates the distinctive contributions made through the analysis of the outputs of research, the structures within which it is organised and the processes by which it is carried out. Quantitative and qualitative data are combined in ways which are sensitive to the contingencies faced at each Centre. The process is necessarily uneven, since some indicators may be inappropriate. The details of the methodology adopted must therefore be tailored to each evaluation, whilst retaining a common approach designed to establish a sound basis of fact and quantitative analysis.

Wellcome Trust (1997). Women and peer review. An audit of the Wellcome Trust decision-making on grants. London: The Wellcome Trust/PRISM (ISBN 1-869835-62-X).

Hochschulevaluation Grossbritannien

Adams, Jonathan, Bailey, Tim, Jackson, Louise, Scott, Peter, Small, Henry & Pendlebury, David (1998). Benchmarking of the International Standing of Research in England: A consultancy study on bibliometric analysis. Centre for Policy Studies in Education at the University of Leeds & Institute for Scientific Information (ISI). http://www.leeds.ac.uk/benchmark/

This study was commissioned by HEFCE to develop indicators of the international standing of English Higher Education Institutions (HEIs) in research across the Units of Assessment (UOAs) established for the 1996 Research Assessment Exercise (RAE). The analysis uses the database of published outputs submitted by university staff to the RAE. Since published outputs were a key indicator by which review panels made their judgments, these publications offer a basis for a comparative measure of the broad international quality of the community. The journal set submitted to a given UOA provides a unique map of publications, determined by the community itself. Mapping ISI's 94 subject categories to the 69 UOAs created a second dataset. There are statistically significant correlations between performance indicators derived from the RAE and ISI data. The RAE dataset included a subset of papers with 'England HEI' addresses. This was used for intra-England analyses, such as England-HEIs vs. England-total. The ISI dataset was used for international comparisons of the relative research performance of England and six other countries: USA, Canada, France, Germany, Australia and Japan. These comparisons included primary analyses of total activity (publications and citations) and secondary indicators of performance (such as citation rates, rebased to world and disciplinary averages as appropriate).

Barnett, Ronald (1996). The evaluation of the higher education system in the United Kingdom. In Robert Cowen (Ed.). World Yearbook of Education 1996: The Evaluation of Higher Education Systems (144-158). London: Kogan Page.

British Council (1998). „Research assessment has damaged British science“ – a debate. Conference Report. Köln: The British Council.

Die große Mehrheit der Seminarteilnehmer stimmte in einem abschließenden Votum gegen die Behauptung, daß die Forschungsevaluierung der britischen Wissenschaft geschadet habe (S. 28).

Burgess, Robert G. (1997). The Peer Review of Teaching and Research in the United Kingdom. In: Herbert Altrichter, Michael Schratz & Hans Pechar (Hrsg.), Hochschulen auf dem Prüfstand (222-234). Innsbruck: Studien-Verlag.

Der Autor stellt anhand der unterschiedlichen Formen externer und interner Begutachtungsverfahren das System der Qualitätssicherung an Großbritanniens Universitäten vor.

Cave, Martin, Hanney, Steve & Henkel, Mary (1995). Performance Measurement in Higher Education - Revisited. Public Money & Management, 15(4), 17-23.

The development of performance indicators (PIs) in higher education in the UK since 1990 is reviewed. The period has witnessed a huge expansion in student numbers, brought to a halt in 1994. The expected development of quantitative indicators has not fully materialized. Instead the focus has shifted to qualitative assessments of teaching and research, which draw in varying degrees upon quantitative data.

Evans, G. R. (1999). Calling Academia to Account: Rights and Responsibilities. Buckingham: Open University Press.

This volume highlights the role of academic freedom in accountability, the role of peer review, procedures for accountability, and related topics.

Geuna, Aldo, Hidayat, Dudi & Martin, Ben (1999). Resource Allocation and Research Performance: The Assessment of Research. A study carried out for the Higher Education Founding Council for England. Brighton: University of Sussex, Science and Technology Policy Research (SPRU).

The 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.

Gillett, R. (1991). Pitfalls in assessing research performance by grant income. Scientometrics, 22(2), 253-263.

The strategy of judging the quality of scientific research by the level of funding it attracts is critically examined. It is argued that an index such as per capita research income, which is based on grant-giver peer review, yields an unsatisfactory measure of scientific performance. It fails to fulfil a basic requirement of a performance indicator, namely, that it should relate outputs to inputs. It has intrinsically low validity, and is strongly confounded with a variety of extraneous factors that are unrelated to research performance.

Glass, Colin J., Hyndman, Noel S. & McKillop, Donal G. (1996). UK Universities: A Time-Series Study of Economies of Scale and Scope in the Context of the Research Assessment Exercise. Public Money & Management, 16(4), 59-64.

A major objective of the Government has been to achieve cost efficiency in the production of teaching and research outputs by UK universities. The article examines recent empirical evidence to support discussions on such issues as the further expansion of the university sector, the targeting of research funding and the desirability of universities producing teaching and research as joint products. The article draws heavily on the two most recent research assessment exercises (RAEs). Furthermore, the particular difficulties and the possible behavioural consequences of using teaching and research as output measures for universities are discussed.

Grant, Jonathan (1999). Evaluating the outcomes of biomedical research on healtcare. Research Evaluation, 8(1), 33-38.

Scientific funding organisations are increasingly being asked to justify their expenditures. For biomedical agencies this ultimately means demonstrating an improvement in healthcare. However, this is particularly problematic because of the complex relationship between research and its incorporation into new treatments. Using clinical guidelines, a new method to follow and quantify the progress of knowledge from biomedical research into clinical practice is proposed. The study demonstrated that the scientific basis of clinical guidelines may be examined using bibliometric techniques of funding source data. This provides a way for funding organisations to measure the impact or output of their funded research on effective, evidence-based medicine.

Henkel, Mary (1999). The modernisation of research evaluation: The case of the UK. Higher Education, 38, 105-122.

The Research Assessment Exercise (RAE), first established by the UK University Grant Committee in 1985 has been an important instrument in the modernisation of higher education in the UK. It is a means of rationalising the stratification of universities and the concentration of research resources, and of maximising research output. At the same time, while its operation

remains substantially under professional control, is has had profound implications for the academic profession.

Henkel, Mary (1998). Evaluation in Higher Education: conceptual and epistemological foundations. European Journal of Education, 33(3), 285-297.

Higher Education Funding Council for England (1999). Future Research Assessment - RAE 2001.
http://www.rae.ac.uk/

Publications

Circulars
RAE 5/99 - Assessment panels' criteria and working methods
RAE 4/99 - Consultation on assessment panels' criteria and working methods
RAE 3/99 - Membership of Assessment Panels
RAE 2/99 - Guidance on Submissions *
RAE 1/99 - Interdisciplinary Research and the RAE *
RAE 4/98 - Chairs of Unit of Assessment Panels *
RAE 3/98 - RAE 2001 and health-related research: consultation *
RAE 2/98 - Research Assessment Exercise 2001: bodies to nominate panel members *
RAE 1/98 - Research Assessment Exercise in 2001: key decisions and issues for further consultation. *
* - on NISS web site

Task Group Reports
Report from the Joint Department of Health / HEFCE task group on Health-related Research and the RAE
Report of the work of the RAE 2001 Task Group on Education Research *
Interim Report of the work of the Joint Funding Bodies/CBI Task Group on User/Industrial Participation in the RAE 2001
* - on NISS web site

Briefing Papers
Interdisciplinary Research and the RAE

Letters
Oct 1998 - Letter requesting Nomination of Panel Members *
* - on NISS web site

Previous RAEs
Publications for previous RAEs can also be found at NISS:
All RAE1996 publications.
All RAE1992 publications.

Higher Education Funding Council for England (1999). Quality assessment reports. Subject-specific reports on the quality of education in HE institutions in England. All are accessible here, although early reports were produced by HEFCE, and later ones by the Quality Assurance Agency (QAA).
http://www.niss.ac.uk/education/hefce/qar/

Higher Education Funding Council for England (1999). Performance Indicators in Higher Education - First Report of the Performance Indicators Steering Group (PISG). Report 99/11, February 1999.
http://www.niss.ac.uk/education/hefce/pub99/99_11.doc

As far as performance indicators are concerned, different stakeholders will regard different indicators as particularly important. The Group's approach has been to develop indicators which will allow stakeholders to extract those which they regard as key and to create their own group of key indicators. The Group commends this approach because of the considerable diversity of the HE sector, both in terms of the missions of institutions, the range of activities undertaken and the nature of their student populations.
Of the wide range of outputs from higher education, the group identified measures of performance relating to: learning and teaching of students, extension of knowledge through research, application of the knowledge and resources of higher education to the needs of business and of society more generally.

Detailed proposals for PIs have been prepared relating to:
a.Widening Participation of under-represented groups
b.Student progression
c.Learning outcomes (including non-completion)
d.Efficiency of learning and teaching
e.Student employment
f.Research output
g.HE links with industry.

Both sector-level and institution-level indicators are proposed under each of these heads, except the last, where only sector-level indicators are proposed.
The group was aware of the need to place such indicators in the context of the institution's circumstances, and to take account of the diversity of the sector. For all the institution-level indicators, therefore, a set of context statistics is provided. In particular, for each relevant indicator concerned with learning and teaching an 'adjusted sector outcome' figure is computed for each institution. This takes account of the intake of students to the institution, their educational backgrounds and the subject mix of that institution. This enables the results for any institution to be compared not with all the other institutions in the sector, but with the average for similar institutions. Such context statistics are designed both to help assess an institution's performance, and to help select comparable institutions with which it is sensible to make comparisons. A list of the indicators and context statistics is set out at Annex E of the report.

Higher Education Funding Council for England (1999). Performance Indicators in Higher Education in the UK. Report 99/66, December 1999.
http://www.niss.ac.uk/education/hefce/pub99/99_66/main.html

The publication of this set of performance indicators is an attempt by the funding councils to provide greater transparency in the way the higher education sector operates. These indicators provide measures of performance with respect to widening access, student progression, outcomes of learning and teaching, learning and teaching efficiency and research output.
Publication follows an extensive consultation with the sector, and more than two years' work by the Performance Indicators Steering Group.
The indicators should be taken as a whole. Non-completion rates for an institution cannot be considered separately from the access indicators, and both should be viewed in the context of the institution's mission. The higher education sector is so diverse that no single measure can adequately describe an institution. However, these indicators are the first step along the road towards providing measures that will reflect this diversity.
The indicators will provide more open information about higher education institutions, provide benchmarks against which performance can be compared, and allow future decisions about the sector to be made on the basis of information that is widely accepted. The funding councils are committed to helping institutions to analyse their data, where this will prove useful.
Because they are mostly based on data from 1997-98, these indicators will not reflect recent activities by institutions, or the impact of funding to recognise the extra costs of recruiting and supporting mature students and those from disadvantaged backgrounds. However, they do provide a baseline against which future indicators can be compared.

Higher Education Funding Council for England (1999). Performance Indicators in Higher Education in the UK: Overview. Report 99/67, December 1999. http://www.niss.ac.uk/education/hefce/pub99/99_67.html

Performance indicators have been introduced for the first time for all 175 publicly-funded universities and colleges in the United Kingdom. This initial set covers: access to higher education, non-completion rates for students, outcomes and efficiencies for learning and teaching in universities and colleges, and research output.
This overview explains the purpose of and background to the exercise, and the national outcomes. It also summarises the content of the six tables of indicators, which encompass a wide range of information about individual institutions.

Higher Education Funding Council for England (1997). Research Assessment Exercise Review: A list of documents which have been published following reviews of the RAE. http://www.niss.ac.uk/education/hefc/rae96/raebrief.html

The following documents have been published following reviews of the RAE. Printed copies are available from the HEFCE.

M 2/97 - Data Collection for the 1996 RAE (Professor Ewan Page) (April). Reports the findings of a study undertaken by Professor Page for the HEFCE.

M 5/97 - Impact of the 1992 RAE on Institutional and Individual Behaviour in English Higher Education: the Evidence from a Research Project (May)

Report on a study conducted for the HEFCE in 1995 and 1996 by Professor Ian McNay of Anglia Polytechnic University. The study provides a range of perceptual and qualitative evidence which includes:
•reports of focus groups of senior academics and administrators
•vignettes of 33 institutions
•staff surveys and surveys of heads of departments
•reports of interviews with some users and funders of research.

M 6/97 - Impact of the 1992 RAE on higher education institutions in England (May)

Draws conclusions on the effects of the 1992 RAE based on evidence from the McNay study and from a study of selectivity conducted for the HEFCE by Segal, Quince and Wicksteed Ltd.

RAE96 1/97 - 1996 RAE Conduct of the Exercise: RAE Manager's Report (May)

Describes how the exercise was conducted and identifies practical issues which may require consideration in the planning of any future, similar exercise.

Higher Education Funding Council for England (1996). 1996 Research Assessment Exercise. http://www.niss.ac.uk/education/hefc/rae96/

Index to papers
You can browse the papers individually (below), or search the collection of available documents.
RAE96 1/97 - Conduct of the Exercise: RAE Manager's Report
RAE96 2/96 - 1996 Research Assessment Exercise: Membership of Assessment Panels (December 96)
RAE96 1/96 - 1996 Research Assessment Exercise: The Outcome (December 96)
RAE96 3/95 - 1996 Research Assessment Exercise: Criteria for Assessment (November 95)
RAE96 2/95 - 1996 Research Assessment Exercise: Guidance on Submissions (November 1995, revised February 1996)

RAE96 1/95 - 1996 Research Assessment Exercise: Membership of Assessment Panels (revised July 96)
RAE96 3/94 - 1996 Research Assessment Exercise: Units of Assessment (November 94)
RAE96 2/94 - Conduct of the Research Assessment Exercise: Panel Membership and Units of Assessment (Consultation) (June 94)

RAE96 1/94 - 1996 Research Assessment Exercise (June 94)

Howells, Jeremy, Nedeva, Maria & Georghiou, Luke (1998). Industry-Academic Links in the UK. A Report to the Higher Education Funding Councils of England, Scotland and Wales. Manchester: University of Manchester: PREST.
http://www.niss.ac.uk/education/hefce/pub98/98_70.html

There has been a spectacular growth in recent years across the United Kingdom in the scale, number and variety of linkages between higher education (HE) and industry. These linkages are manifested in research collaboration, provision of consultancy services, market transactions in the commercialisation of research, and industry's growing involvement as an interactive user of all types of teaching and training. Through surveys of industrial liaison officers and continuing education officers, interviews with senior staff, and compilation of available statistics, this report describes the status and trends of these relationships.

Johnes, Geraint (1994). Research Performance Measurement: what can international comparisons teach us? Comparative Education, 30(3), 205-216.

The current systems of financing research in each of three countries - Australia, Canada and the United Kingdom - are described and the three countries are used to illustrate a number of problems which need to be addressed in tackling the problem of research evaluation and the link with public policy. Particular attention is paid to three such problems: the link between evaluation and funding; issues of incentive compatibility; and the impact which performance evaluation can have on the behaviour of those whose performance is being evaluated.

Johnes, Jill (1996). Performance assessment in higher education in Britain. European Journal of Operational Research, 89, 18-33.

A methodology is developed in the framework of production theory and uses multiple regression techniques to estimate the relationship between the outputs and inputs of universities. Around 80% of the inter-university variation in four output measures can be explained by corresponding variations in several input measures. This highlights the need to take into account the inputs available to a university when comparing its output performance with that achieved by other institutions. The problems of interpreting an array of performance indicators are also clearly demonstrated.

May, Robert M. & Sarson, Stuart C. (1999). Revealing the hidden costs of research. Nature, 398 (8 April), 457-459.

How should universities account for the money they receive from government? The answer is not as simple as it may at first appear. There are valuable lessons that other countries can learn from the US experience.

May, Robert M. (1998). The Scientific Investments of Nations. Science, 281, 49-51.

May, Robert M. (1997). The Scientific Wealth of Nations. Science, 275, 793-796.

McNay, Ian (1999). The paradoxes of research assessment and funding. In Mary Henkel & B. Little (Eds.), Changing Relationships between Higher Education and the State. London: Jessica Kingsley Publishers.

Nuttall, W. J. (1999). Engineering and Physical Sciences Research Council (UK) peer-review procedures 1995-1996. Research Evaluation, 8(2), 132-140.

The paper reports on a survey of UK members of the Institute of Physics selected on the basis of an expected research association to the Engineering and Physical Sciences Research Council (EPSRC) during the years 1995 and 1996. The survey assesses the EPSRC's 'college' system for research evaluation from the perspective of members of the EPSRC's colleges and from those researchers who have only seen the system from the outside. Insights into the workings of a college-based peer-review system are gained, particularly in regard to referee expertise and the most effective provision of information to prioritisation committees who must award or decline applications formally based on referee opion of an applicant's submission.

Patel, Pari & Pavitt, Keith (1995). Patterns of Technology Activity: their Measurement and Interpretation. In Paul Stoneman (Ed.), Handbook of the Economics of Innovation and Technological Change (14-51). Oxford: Blackwell.

The authors address fundamental issues relating to the measurement and interpretation of patterns of technological activity. They discuss R&D, patenting activity, and several other measures of technological activity illustrating the availability of data and the patterns to be observed in that data as regards international, national, industrial and firm level technological performance and changes therein over time. They also discuss the role of transnational corporations in the technology process, such corporations considerably complicating any supposed simple relationship between national technological activity and national technological performance.

Patrick, William J. & Stanley, Elizabeth C. (1998). Teaching and Research Quality Indicators and the Shaping of Higher Education. Research in Higher Education, 39(1), 19-41.

Two important sets of performance indicators have become established in the United Kingdom: research quality ratings and teaching quality ratings. The research quality ratings and, to a lesser extent, the teaching quality ratings, influence the level of government funding provided to higher education institutions. This paper considers the correlation between the two ratings and the possible consequences of policies that reshape the higher education sector by concentrating research resources in a limited number of institutions. Comparisons are made between quality assurance/assessment approaches in the United Kingdom and those in the United States.

Patrick, William J. & Stanley, Elizabeth C. (1996). Assessment of Research Quality. Research in Higher Education, 37(1), 23-42.

The paper describes the British experience of nation-wide research quality assessment exercises, and newly introduced measures intended to improve accountability.

Pavitt, Keith (1998). Do patents reflect the useful research output of universities? Research Evaluation, 7(2), 105-111.

As the costs of data handling continue to decrease, patenting information is being used increasingly by analysts and practitioners to deepen their understanding of the nature, sources and consequences of technical change. On the basis of earlier analyses, we conclude that patents granted to universities give a very partial and distorted picture of the contributions of university research to technical change. However, citations in patents to published research papers, together with collaborative publications between universities and industry, offer rich and rewarding sources of information on how university research contributes to technical change.

Salter, Ammon J. & Martin, Ben R. (1999). The Economic Benefits of Publicly Funded Basic Research: A Critical review. Brighton: University of Sussex, Science and Technology Policy Research (SPRU). Paper No. 34.
http://www.sussex.ac.uk/spru/docs/sewps/sewp34/sewp34.html
http://www.sussex.ac.uk/spru/docs/sewps/index.html

This article critically reviews the literature on the economic benefits of publicly funded basic research. In that literature, three main methodological approaches have been adopted - econometric studies, surveys and case studies. Econometric studies are subject to certain methodological limitations but they suggest that the economic benefits are very substantial. These studies have also highlighted the importance of spillovers and the existence of localisation effects in research. From the literature based on surveys and on case studies, it is clear that the benefits from public investment in basic research can take a variety of forms. We classify these into six main categories, reviewing the evidence on the nature and extent of each type. The relative importance of these different forms of benefit apparently varies with scientific field, technology and industrial sector. Consequently, no simple model of the economic benefits from basic research is possible. We reconsider the rationale for government funding of basic research, arguing that the traditional 'market failure' justification needs to be extended to take account of these different forms of benefit from basic research. The article concludes by identifying some of the policy implications which follow from this review.

The Times (1999). The Times Good University Guide. London: Times Books. http://www.the-times.co.uk/gug/

The Times of London offers its latest ranking of higher education institutions in the United Kingdom. Universities are rated as a whole and various subject departments are also rated. The latest overall rankings are presented in a league table (i.e., comparative data table) in the April 23, 1999 issue of The Times. Links to past years' surveys are also provided.

Webster, Andrew (1994). International evaluation of academic-industry relations: contexts and analysis. Science and Public Policy, 21(2), 72-78.

An analytical framework has been developed which can be used to explore and evaluate the role of academic-industry collaborations in the wider innovation system.

Williams, Ruth (1997). Quality Assurance and Diversity: The Case of England. In John Brennan, Peter de Vries & Ruth Williams (Eds.), Standards and Quality in Higher Education (104-118). London: Jessica Kingsley Publishers.

The chapter describes the arrangements for quality assurance - quality assessment as conducted by the Higher Education Funding Council for England (Internet-Adresse: http://www.hefce.ac.uk/) and quality audit by the Higher Education Quality Council (Internet-Adresse: http://www.niss.ac.uk/education/heqc/. The Higher Education Quality Council has been replaced by the Quality Assurance Agency. You should see their web pages for current information: http://www.qaa.ac.uk/) - and the outcomes and issues emerging from these two processes. In contrast to quality assessment whose focus is at the subject level, quality audit, undertaken by HEQC, is concerned with institional systems and procedures which assure quality and standards.

5. Deutschland

Forschungsevaluation

Arieh, Asaf Ben, Grupp, Hariolf & Maital, Shlomo (1998). Optimal incremental innovation: an evaluative approach for integrating R&D and marketing.

New operational definitions of incremental innovation, standard innovation, and radical innovation, are constructed using the Fh-ISI 'technometric benchmarking' model. Based on this definition, optimal incremental innovation is formulated as a linear programming problem. The model is illustrated by an actual case: reconfiguration of a gamma camera. It is shown how this model can contribute to improved allocation of research and development (R&D) resources, by integrating marketing and R&D in a single decision-support model.

Becher, Gerhard & Kuhlmann, Stefan (Eds.). (1995). Evaluation of Technology Policy Programmes in Germany. Dordrecht: Kluwer Academic Publishers.

The book attempts to do two things. First, it presents the state of the art in evaluation methodology in Germany, by bringing together a number of authors belonging to the Study Group on Technology Policy Evaluation Research, from some of the top research institutes in Germany and Switzerland. Second, it is a call for a critical international debate on evaluation experiences and methodologies.

Beise, Marian & Stahl, Harald (1998). Public Research and Industrial Innovations in Germany. Mannheim: Zentrum für Europäische Wirtschaftsforschung GmbH (Discussion Paper No. 98-37).

Wie viele industrielle Innovationen bringt die öffentliche Forschung hervor? Die vorliegende Untersuchung versucht, diese Frage zu beantworten und damit die wirtschaftliche Rechtfertigung für Finanzierung und Durchführung von natur- und ingenieurwissenschaftlicher Forschung in öffentlichen Forschungseinrichtungen empirisch zu überprüfen. Hierfür wurden rund 2.300 Unternehmen in der vierten Innovationserhebung von ZEW und infas im Jahr 1996 gefragt, ob sich unter ihren zwischen 1993 und 1995 eingeführten Innovationen neue Produkte und Prozesse befanden, die ohne die neueren Forschungsergebnisse von Hochschulen und anderen öffentlich finanzierten Forschungseinrichtungen nicht oder nur mit zeitlicher Verzögerung von mehr als einem Jahr aufgenommen worden wären. Knapp 9 % der Unternehmen mit Innovationen bejahten diese Frage. Mit den Produkten, die nur mit Hilfe der öffentlichen Forschung zustande kamen, wurden 1995 allerdings weniger als 5 % des gesamten Umsatzes mit neuen Produkten gemacht. Der Technologietransfer von den öffentlichen Forschungseinrichtungen wird dabei entscheidend von den eigenen FuE-Aktivitäten der Unternehmen begünstigt. Sie sind nötig, um die wissenschaftlichen Ergebnisse im Unternehmen in marktgerechte Innovationen umzusetzen. Dagegen wirkt sich die räumliche Nähe zu Forschungseinrichtungen, die unter dem Stichwort regionale High-Tech-Agglomerationen diskutiert wird, nicht auf den Technologietransfer aus.

Boehmer, Alexander von (1995). Internationalisierung industrieller Forschung und Entwicklung – Typen, Bestimmungsgründe und Erfolgsbedingungen. Wiesbaden: Deutscher Universitätsverlag (Betriebswirtschaftslehre für Technologie und Innovation, Band 13).

Auf der Grundlage einer umfangreichen Fragebogenerhebung bei multinationalen Unternehmen in der Bundesrepublik Deutschland, den USA und Großbritannien wird ein differenziertes Bild ausländischer F&E-Aktivitäten gewonnen.

Buchholz, K. (1995). Criteria for the analysis of scientific quality. Scientometrics, 32(2), 195-218.

A systematic compilation of criteria which covers the full range of excellence to failure with respect to scientific quality is developed and a comprehensive list of criteria is presented which should provide a basis both for objective and adequate characterization of publications.

Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (1997). Förderung von Erfindungen und Patenten im Forschungsbereich. Bonn: BLK (Eine Online-Fassung des Heftes kann eingesehen werden unter: http://www.patente.bmbf.de/patent/inhalt11.htm).

Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (1998). Heft 61: Beschluß der Regierungschefs „Sicherung der Qualität der Forschung“. Bonn: BLK.

Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (1999). Forschungsförderung in Deutschland. Bericht der internationalen Kommission zur Systemevaluation der Deutschen Forschungsgemeinschaft und der Max-Planck-Gesellschaft. Bonn: BLK.

Anlaß für die Systemevaluation der gemeinsam geförderten Forschungseinrichtungen ist ein Beschluß der Regierungschefs von Bund und Ländern vom Dezember 1996. Als erster Bericht wurden Empfehlungen zur Fraunhofer-Gesellschaft im Februar 1999 vorgelegt. Für DFG und MPG wurde die Einbindung in die deutsche Forschungslandschaft untersucht und geprüft, welche Empfehlungen sich daraus für Bund und Länder ableiten lassen.
Der Vorsitzende der Evaluierungskommission, Richard Brook, bescheinigte der DFG und der MPG eine führende Rolle in der deutschen Forschung. Beide Einrichtungen arbeiteten auf hohem internationalem Niveau und seien wesentliche Eckpfeiler für das Forschungssystem in Deutschland und Europa. Die Kommission empfiehlt eine größere Beweglichkeit und Durchlässigkeit im Forschungssystem, insbesondere im Hinblick auf die Verbindungen in die Wirtschaft. Darüber hinaus seien die Profilbildung der Universitäten zu steigern und neue Akzente in der strategischen Orientierung von DFG und MPG zu setzen. Besonderen Wert legt die Kommission auf die möglichst frühe Selbständigkeit des wissenschaftlichen Nachwuchses und plädiert für einen Verzicht auf die Habilitation. Auch solle die starke disziplinäre Orientierung an den Universitäten zugunsten von beweglichen Organisationsformen für die temporäre Zusammenarbeit verschiedener Disziplinen gelockert werden. Die Kommission hebt hervor, daß viele Universitäten und Fachbereiche qualitativ hochwertige, international anerkannte und wettbewerbsfähige Leistungen in der Forschung erbringen. Sie weist jedoch auch darauf hin, daß die Leistungsfähigkeit der deutschen Wissenschaft durch Verfassung und Strukturen der Universitäten erheblich eingeschränkt werde.
Die Kommission regt ferner an, bei der DFG die Gutachterstruktur zu ändern und eine wissenschaftsstrategische Programmsteuerung zu entwickeln sowie die Struktur der Förderinstrumente zu überprüfen.
Bei der MPG wird unter anderem eine stärkere Zusammenarbeit mit den Universitäten empfohlen. Darüber hinaus wird angeregt, das Prinzip der Förderung in Form von Instituten zu überdenken und um flexible, zeitlich begrenzte Arbeitsformen und Förderinstrumente zu ergänzen. Die strategische Planung sei durch die Einsetzung eines extern besetzten Beratungsgremiums des Präsidenten zu stärken. Auch die Berufungsverfahren der MPG sollten vereinfacht und deutlich beschleunigt werden und eine Verschlankung der Generalverwaltung angestrebt werden.
Der Text der gemeinsamen Stellungnahme von DFD und MPG kann der folgenden WWW-Adresse entnommen werden: http://www.dfg.de/aktuell/pressemitteilungen/forschungspolitik/presse_1999_57.html
Der Test der Stellungnahme der DFG findet sich unter der folgenden WWW-Adresse: http://www.dfg.de/aktuell/download/evaluation.html

Cohausz Hannig Dawidowicz & Partner (1999). Untersuchung zum Verwertungsprivileg – Relevanz des sog. Hochschullehrerprivilegs nach § 42 ArbNErfG. Bonn: Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie.

Die vorliegende Studie ist mit dem Ziel vergeben worden, die Relevanz des Verwertungsprivilegs der Hochschullehrer für das praktische Patentgeschehen im Hochschulbereich und die möglichen oder wahrscheinlichen Folgen einer Änderung zu untersuchen.

Creutzfeldt, Werner & Gerok, Wolfgang (Hrsg.). (1997). Medizinische Publizistik: Probleme und Zukunft. Stutttgart: Georg Thieme Verlag.

Der Anhang enthält u. a. Tabellen und Statistiken über die Rangordnung von biomedizinischen Zeitschriften sowie deutschen medizinischen Fakultäten und ihren Mitgliedern.

Ernst, Holger (1996). Patentinformationen für die strategische Planung von Forschung und Entwicklung. Wiesbaden: Deutscher Universitäts-Verlag.

Beschrieben werden eine Vielzahl von Anwendungsmöglichkeiten strategischer Patentanalysen: Technologische Konkurrenzanalyse; FuE-Mangement; Externe Technologiebeschaffung; Patent-Portfolio-Management; Marktüberwachung; FuE-Personalmanagement.

FhG (Hg.). (1998). Systemevaluierung der Fraunhofer-Gesellschaft. Bericht der Evaluierungskommission. München.

Fleischer, Manfred (1999). Innovation, Patenting and Performance. Economie Appliquée, 52(2), 95-119.

Am Beispiel einer wichtigen Investitionsgüterindustrie wird der Einfluß nationaler Patentsysteme auf den Unternehmenserfolg untersucht. Zahlreiche Theorien gehen davon aus, daß das Innovationsverhalten von Unternehmen von nationalen und sektoralen Innovationssystemen beeinflußt wird. Das Patentsystem spielt in diesen Systemen eine wichtige Rolle. Auf der Unternehmensebene beeinflußt es die Möglichkeiten der Unternehmen, Gewinne zu erzielen. Am Beispiel der 49 größten Unternehmen der internationalen Werkzeugmaschinenindustrie wird nachgewiesen, daß in diesem Industriezweig Patente ein wichtiges Instrument zur Erzielung von Gewinnen darstellen. Dies zeigt sich besonders deutlich an der engen positiven Beziehung, die zwischen dem Marktwert der Unternehmen und ihrem Patentierverhalten empirisch festgestellt wurde.

Frömmel, Cornelius & Heß, Dieter (Hrsg.). (1998). Leistungsbewertung Forschung: Förderung der Forschung an Medizinischen Fakultäten und Hochschulen. Berlin: Klarsicht Verlag. ISBN 3-930474-02-6

Glänzel, Wolfgang (1996). The need for standards in bibliometric research and technology. Scientometrics, 35(2), 167-176.

The need for standardisation in bibliometric research and technology is discussed in the context of failing communication within the scientific community, the unsatisfactory impact of bibliometric research outside the community, and the observed incompatibility of bibliometric indicators produced by different institutes.

Grupp, Hariolf (1997). Evaluation of Research and Development Programs by Technology Indicators. In Mark S. Frankel & Jane Cave (Eds.), Evaluating Science and Scientists: An East-West Dialogue on Research Evaluation in Post-Communist Europe (183-201). Budapest: Central European University Press.

After identifying four main objectives of program evaluation (assessment of program assumptions, of goal achievements, of impacts, and of program administration), the author discusses the use of patents and other technology indicators ('technometrics') in the evaluation of applied research

programs. He uses two cases - the development of lasers for metal work by a program supported by the European Community, and photovoltaic R&D in Germany - to demonstrate how various technology indicators can be used to evaluate applied research programs.

Herbertz, Heinrich & Müller-Hill, Benno (1995). Quality and efficiency of basic research in molecular biology: a bibliometric analysis of thirteen excellent research institutes. Research Policy, 24, 959-979.

The authors try to assess the research performance of 13 research institutes active in the field of molecular biology. For this purpose they have counted the number of scientific publications and the number of citations received during a five-year period. They use citations per publication as an indicator of quality and costs per citation as an indicator of efficiency of research. Peer review seems to discourage uninterested, i.a. not cited, research. Grant systems seem to work more efficiently than funding on a permanent institutional basis.

Heuer, Herbert, Fuhrmann, Hartwig & Schmidt, Klaus-Helmut (1998). Die Beurteilung von Forschungsleistungen. Das Beispiel des Instituts für Arbeitsphysiologie an der Universität Dortmund. Frankfurt am Main: Peter Lang.

In den letzten Jahren läßt sich eine zunehmende Tendenz beobachten, Forschungsleistungen in systematischer Weise zu bewerten. Die Schwierigkeiten dabei haben wesentlich mit der Existenz unterschiedlicher Fachkulturen zu tun sowie mit den verschiedenartigen Aufgaben und Zielsetzungen unterschiedlicher Forschungseinrichtungen. Die Versuchung ist groß, auf leicht verfügbare quantitative Indikatoren zurückzugreifen, wie etwa die Anzahl begutachteter Publikationen bzw. die Summe der Impact-Faktoren und die eingeworbenen Drittmittel. Ein solches Vorgehen wird der vorhandenen Vielfalt von Forschungsaufgaben nicht gerecht und führt langfristig zu ihrer Einengung. Berichtet wird von der Übertragung eines in der Industrie bewährten Verfahrens auf eine Forschungseinrichtung, das der Vielfältigkeit der Aufgaben verschiedener Forschungseinrichtungen Rechnung trägt.

HGF (Hg.). (1997). Begutachtungen in der Helmholtz-Gemeinschaft. Grundsätze und Verfahren. Dokumentation 15. Bonn.

Hornbostel, Stefan (1997). Wissenschaftsindikatoren – Bewertungen in der Wissenschaft. Opladen: Westdeutscher Verlag.

Wissenschaft hat nach Hornbostel den Charme einer liebenswürdig chaotischen Gelehrtenstube verloren und ist zur „big science“ geworden. Damit ist sowohl ein neuartiger Reflexionsbedarf entstanden als auch eine Fülle von Steuerungsproblemen. Wissenschaftsindikatoren sind eine Antwort auf die veränderten Rahmenbedingungen. Sie sollen Strukturen, Quantitäten und Qualitäten wissenschaftlicher Produktion messbar machen. Ihre Entwicklung verlief jedoch in einem sehr pragmatischen Kontext, mit der Folge, dass der Kontakt zur Wissenschaftstheorie verloren ging und die Beurteilung des Nutzens von Indikatoren nicht selten in einen Glaubenskrieg ausartete. Mit dem Band „Wissenschaftsindikatoren“ wird nicht nur eine Verbindung zwischen Wissenschaftstheorie, -soziologie und der Indikatorenforschung geschaffen, sondern auch eine detaillierte Darstellung der Leistungsfähigkeit und der Grenzen von Wissenschaftsindikatoren vorgelegt.

Kuhlmann, Stefan (1999). Distributed Intelligence for Innovation Policy Planning: Integrating Evaluation, Foresight and Technology Assessment. In: Susanne Bührer & Stefan Kuhlmann (Eds.), Evaluation of Science and Technology in the new Europe. Proceedings of an International Conference on 7 and 8 June 1999, Berlin (137-145). Bonn / Bruxelles: Federal Ministry of Education and Research / European Commission.

Kuhlmann, Stefan (1999). Evaluation of Research Policy as a Moderation Process. In Vaclav Paces, Ladislav Pivec & Albert H. Teich (eds.), Science Evaluatiion and Its Management (8-16). Ohmsha: IOS Press, NATO Science Series: Science & Technolgy Policy, vol. 28.

Evaluation procedures measure the scientific and technological quality or the soio-economic impact of publicly funded research. But why couldn't evaluation procedures also be used to 'moderate' struggles, controversies and negotiations in the science and technology policy arena? The book chapter addresses this question, utilising the German evaluation method which is characterised by a relatively high degree of institutional differentiation and autonomy of the major policy actors. After some theoretical considerations, a case study is presented illustrating the 'moderation approach', i. e., a multi-annual monitoring evaluation of eight newly created, publicly funded interdisciplinary clinical research centres at German university hospitals.

Kuhlmann, Stefan (1998). Politikmoderation. Evaluationsverfahren in der Forschungs- und Technologiepolitik. Baden-Baden: Nomos. (ISBN 3-7890-5534-4)

Kuhlmann, Stefan (1998). Moderation of Policy-Making. Science and Technology Policy Evaluation Beyond Impact Measurement - The Case of Germany. Evaluation, 4(2), 130-148.

Kuhlmann, Stefan (1997). Evaluation as a Medium of Science and Technology Policy: Recent Developments in Germany and Beyond. In OECD (Ed.), Policy Evaluation in Innovation and Technology Towards Best Practices (Chapter 25). Paris: OECD. http://www.oecd.org/dsti/sti/stat-ana/prod/evaluation.htm

Kuhlmann, Stefan (1995). Patterns of science and technology policy evaluation in Germany. Research Evaluation, 5(1), 23-33.

In Germany the practice of the evaluation of research and of research institutions is characterised by a considerable degree of self-organisation of the scientific community. A report is given of a recent comprehensive analysis of evaluation practice in S&T programmes: it critically analysed over 50 evaluation studies which the Federal Ministry for Research and Technology had commissioned since 1985. On the basis of this analysis and its recommendations, a rough outline for a systematised future evaluation practice was proposed. Recent attempts to combine strategically oriented evaluations of project funding with those of institutional funding of S&T organisations are reported. In this context, several studies of future developments of science and technology ('technology foresight') were recently carried out, raising some public interest.

Kuhlmann, Stefan (1995). German government department's experience of RT&D programme evaluation and methodology. Scientometrics, 34(3), 461-471.

Kuhlmann, Stefan & Holland, Doris (1995). Evaluation von Technologiepolitik in Deutschland. Konzepte, Anwendung, Perspektiven. Heidelberg: Physica-Verlag.

Die staatliche Forschungs- und Technologiepolitik braucht transparente Planungs- und Bewertungsprozesse, damit der erforderliche Dialog der Akteure offen und informiert geführt werden kann. Das Buch dokumentiert und analysiert 50 Evaluationsstudien, die das Bundesministerium für Forschung und Technologie seit 1985 hat durchführen lassen. Es diskutiert auf dieser Grundlage die inhaltlichen und methodischen Konzepte, die bisherigen praktischen Erfahrungen und die Nutzungs- und Entwicklungspotentiale solcher Evaluationsstudien. Das Buch enthält außerdem umfangreiche und detaillierte Empfehlungen für die künftige Evaluationspraxis im Bereich der Forschungs- und Technologiepolitik.

Krull, Wilhelm (1999). Publications, Pressure to Publish, Selection Procedures of Scientific Journals, Citation Indices as Factors Producing Scientific Misconduct? Paper presented at the Ringberg Conference on „Ethics in Research“, 20-23 October 1999. München: Max-Planck-Gesellschaft.

Krull, Wilhelm (1995). The Max Planck Experience of Evaluation. Scientometrics, 34(3), 441-450.

The paper outlines the various ways and levels of quality assessment within the Max Planck system. In particular, it emphasizes the importance of ex ante-evaluation, and the need for an assessment of ongoing research work at regular intervals. Furthermore, the strengths and weaknesses of quantitative indicators are discussed, and, finally, some principles for policy-relevant evaluations formulated.

Krull, Wilhelm & Winter, Ekkehard (1996). Reader zum Seminar „Von der Evaluation zur Prospektion? – Neue Herausforderungen für die deutsche Wissenschafts- und Forschungspolitik“ (München, 10. – 12. November 1995). Essen: Stifterverband für die Deutsche Wissenschaft.

Laudel, Grit (1999). Interdisziplinäre Forschungskooperation: Erfolgsbedingungen der Institution „Sonderforschungsbereich“. Berlin: edition sigma.

Obwohl die Einrichtung eines SFB mit außerordentlich hohen Transaktionskosten verbunden ist - es sind mindestens einjährige Anstrengungen einer Gruppe von Wissenschaftlern erforderlich -, überwiegt im Urteil der Wissenschaftler die positive Bewertung des Förderinstruments SFB. Die Studie zeigt ferner, daß das SFB-Programm die beabsichtigte kooperationsfördernde Wirkung erzielt.

Maurer, Michael (1996). Evaluation von Forschungsinstituten und -feldern durch den Wissenschaftsrat. In Wilhelm Krull & Ekkehard Winter (Hrsg.), Reader zum Seminar „Von der Evaluation zur Prospektion? – Neue Herausforderungen für die deutsche Wissenschafts- und Forschungspolitik“ (München, 10. – 12. November 1995). Essen: Stifterverband für die Deutsche Wissenschaft.

Max-Planck-Gesellschaft (1998). Regelungen für das Fachbeiratswesen. München: MPG.

Das Fachbeiratswesen ist das zentrale Element der begleitenden Evaluation der Forschungseinrichtungen der MPG. Eine regelmäßige Evaluation ihrer Institute liegt im Interesse der MPG und trägt zur Funktionsfähigkeit ihres Selbststeuerungssystems bei; sie dient gegenüber der Öffentlichkeit zur Rechenschaftslegung über den sinnvollen und effektiven Einsatz der ihr zur Verfügung gestellten Mittel. Die Regelungen für das Fachbeiratswesen sollen allen Beteiligten als verbindlicher Leitfaden für das Vorgehen bei der Begutachtung der Institute durch die Fachbeiräte dienen.

Max-Planck-Gesellschaft (1999). Statusbericht für den Fachbeirat: notwendiger Inhalt. München: MPG.

Der Statusbericht des Instituts - der eine umfassende Selbstdarstellung der wissenschaftlichen Leistungsfähigkeit und der Ausstattung des Instituts und seiner Abteilungen bzw. Arbeitsbereiche beinhalten soll - ist die wesentliche schriftliche Grundlange für die Arbeit des Fachbeirats. Die nachstehend genannten Punkte sollte der Statusbericht - sowohl für die regelmäßige Begutachtung alle zwei Jahre als auch für die erweiterte Evaluation alle sechs Jahre - in jedem Fall beinhalten.

Meyer-Krahmer, Frieder & Schmoch, U. (1998). Science-based technologies: university-industry interactions in four fields. Research Policy, 27, 835-851.

Meyer-Krahmer, Frieder & Reiss, Thomas (1992). Ex ante evaluation and technology assessment - two emerging elements of technology policy evaluation. Research Evaluation, 2(1), 47-54.

Ex ante evaluations of government programmes promoting R&D are rare: this is a serious deficit since no systematic basis for decision making is formed. The article describes concepts and methods of programme evaluations, as well as the strengths and weaknesses of evaluation research.

The technometrics approach is presented as a method of evaluation of biotechnology R&D which contributes to ex ante evaluations. Finally, technology assessment can be combined with programme evaluation.

Röbbecke, Martina & Simon, Dagmar (1999). Zwischen Reputation und Markt – Ziele, Verfahren und Instrumente von (Selbst)Evaluationen außeruniversitärer, öffentlicher Forschungseinrichtungen. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (P 99 – 002).

In den vergangenen Jahren wurden die wissenschaftspolitischen Diskussionen in erheblichem Maße von der Frage bestimmt, wie – insbesondere angesichts stagnierender oder sogar zurückgehender Ressourcen – die Qualität der Forschung gesichert werden kann. Zunächst konzentrierten sich die Bemühungen zur Qualitätssicherung wesentlich auf die Hochschulen; in jüngster Zeit rücken auch die außeruniversitären Forschungseinrichungen – insbesondere die zur Zeit durch den Wissenschaftsrat evaluierten Einrichtungen der Blauen Liste – in das Zentrum der Aufmerksamkeit. Röbbecke und Simon setzen sich mit Zielen, Verfahren und Instrumenten von Evaluationen der in ihren Aufgabenstellungen und Zielen ausgesprochen heterogenen Forschungsinstitute der Wissenschaftsgemeinschaft G. W. Leibniz (WGL) auseinander. Im Hinblick auf die Heterogenität gewinnt die Frage nach den adäquaten Bewertungsmaßstäben besondere Relevanz. Dabei geht es zunächst um die Suche nach gültigen Indikatoren – ausgehend von der Annahme, dass die gängigen Wissenschaftsindikatoren das Leistungsspektrum der WGL-Institute – die Verbindung von Grundlagen- und anwendungsorientierter Forschung sowie die Beratungs- und Dienstleistungen – nur unzureichend erfassen. In einem zweiten Schritt gehen die Autorinnen auf die Ziele von Evaluationen ein und es werden Elemente eines noch weiterzuentwickelnden – auf Qualitätsförderung ausgerichteten – Modells vorgestellt, das interne Selbstbeobachtungs- und externe Begutachtungsprozesse integrieren soll. Ferner gehen die Autorinnen der Hypothese nach, dass die Leistungsfähigkeit einer Forschungseinrichtung entscheidend damit zusammenhängt, welche organisatorischen Lösungen für die Durchführung komplexer Forschungs- und Beratungsaufgaben gefunden werden. Röbbecke und Simon plädieren dafür, dass neben den Forschungsergebnissen die Forschungsorganisation – beispielsweise interne Steuerungs- und Managementstrukturen – einen wichtigen Gegenstand von Evaluationen darstellt.

Röbbecke, Martina & Simon, Dagmar (Hrsg.). (1999). Qualitätsförderung durch Evaluation? Ziele, Aufgaben und Verfahren von Forschungsbewertungen im Wandel. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (P 99 – 003).

Das Paper dokumentiert einen Workshop, der sich mit Evaluationen außeruniversitärer, staatlich finanzierter Forschungseinrichtungen auseinandergesetzt hat. Im Unterschied zu den Hochschuldebatten über Evaluationen, die bereits in den achtziger Jahren eingesetzt haben und die sich nicht nur in einer Fülle von Publikationen, sondern auch in einer ansehnlichen Zahl von Reformprojekten niederschlagen, finden im außeruniversitären Sektor Diskussionen in einem vergleichbaren Maß nicht statt. Dabei können viele Forschungseinrichtungen auf langjährige Evaluationserfahrungen zurückblicken: sei es auf interne Bewertungsverfahren durch Institutsbeiräte, sei es auf externe Evaluierungen durch denWissenschaftsrat. Mit dem Workshop war intendiert, eine Debatte über Ziele, Aufgaben, Verfahren und Instrumente von Forschungsevaluationen aus den unterschiedlichen Perspektiven von Wissenschaftspolitikern, Evaluationsspezialisten sowie Praktikern aus Hochschulen und außeruniversitären Forschungseinrichtungen zu führen.
In den Beiträgen werden die Erwartungen an Evaluationen unter veränderten wissenschaftspolitischen Rahmenbedingungen diskutiert sowie die Schwierigkeiten erörtert, die Qualität wissenschaftlicher Leistungen „festzumachen“. Darüber hinaus wird die Frage aufgeworfen, welche Bewertungskriterien für die Evaluation heterogener Institutstypen angemessen sind, und die Bedeutung von Selbstevaluationen als Instrument der Qualitätssicherung und –förderung herausgestellt. Nach der Präsentation des niederländischen Evaluationsansatzes, der erheblichen Einfluß auf die bundesdeutsche Diskussion genommen hat, resümiert ein abschließender Beitrag Entwicklungen der Hochschulevaluationen im internationalen Vergleich.

Simon, Dagmar & Röbbecke, Martina (1999). Self-evaluation as a controlling instrument. In Assessing Assessments - European Experiences (S. 62-66). Proceedings of a conference organized by the Danish Institute for Studies in Research and Research Policy in cooperation with The European Consortium for Political Research. Aarhus: The Danish Institute for Studies in Research and Research Policy.

Tegelbekkers, Friedrich (1997). Evaluation of the Blue List Institutes by the Science Council in Germany. In OECD (Ed.), The Evaluation of Scientific Research: Selected Experiences (83-90). Paris: OECD/GD(97)194.

Vitt, Jan (1999). Schlüsselerfinder und Nachwuchsstars. Wissenschaftsmanagement, 5(5), 30-33.

In insgesamt 43 deutschen Industrieunternehmen aus der chemischen Industrie, der Elektrotechnik und dem Maschinenbau wurde die Verteilung erfinderischer Leistung untersucht. Zur Leistungsmessung wurde einerseits die Anzahl der von einem Erfinder hervorgebrachten Patente ausgewertet. Andererseits wurden auch Qualitätsaspekte berücksichtigt, indem Patente hinsichtlich ihres Einflusses auf den technologischen und ökonomischen Unternehmenserfolg qualitativ gewichtet wurden. Beispielsweise haben Patente, die zusätzlich bei einem ausländischen Patentamt angemeldet worden sind, einen hohen ökonomischen Wert. Weitere Qualitätsmerkmale eines Patentes sind dessen Erteilung durch die zuständige Patentbehörde, die Aufrechterhaltung der Gültigkeit und die Erwähnung auf anderen Patenten (Patentzitat).

Vitt, Jan (1998). Schlüsselerfinder in der industriellen Forschung und Entwicklung. Wiesbaden: Deutscher Universitätsverlag.

Wagner-Döbler, R. (1994). The frequency distribution of legal decision citations in the German jurisdiction. Scientometrics, 29(1), 15-26.

Wissenschaftliche Kommission Niedersachsen (1999). Forschungsevaluation an niedersächsischen Hochschulen und Forschungseinrichtungen - Grundzüge des Verfahrens. Hannover.

Wissenschaftsrat

Fach- und institutionenübergreifende Analyse und Bewertung großer Forschungsfelder durch den Wissenschaftsrat:
Mit seiner im Jahr 1994 verabschiedeten Stellungnahme zur Umweltforschung in Deutschland hatte der Wissenschaftsrat zum ersten Mal eine Querschnittsstudie zu einem großen Forschungsfeld vorgelegt, das über die Grenzen der klassischen Fachdisziplinen und der Sektoren der öffentlich finanzierten Forschungseinrichtungen hinweg bewertet wurde. Für die sich anschließende Querschnittsbegutachtung der Materialforschung wurde ein zweistufiges Verfahres gewählt. Zunächst wurden „Empfehlungen zur Förderung materialwissenschaftlicher Forschung und Lehre an den Universitäten“ vorgelegt, die sich vor allem mit den Strukturen materialwissenschaftlicher Forschung und Lehre an den Universitäten in Deutschland befassen. Darauf aufbauend wurde die „Stellungnahme zur außeruniversitären Materialwissenschaft“ ausgearbeitet, in der mehr als 30 Einrichtungen der Materialforschung dargestellt und bewertet wurden. Der Wissenschaftsrat sieht in der fach- und institutionenübergreifenden Analyse und Bewertung großer Forschungsfelder eine wichtige Aufgabe zum Erhalt und zur Steigerung der Leistungsfähigkeit der Forschung in Deutschland. Mit der Stellungnahme zur Energieforschung werden die Querschnittsbewertungen fortgesetzt, wobei universitäre und außeruniversitäre Einrichtungen gemeinsam betrachtet werden.

Wissenschaftsrat (1999). Stellungnahme zur Energieforschung. Köln.

Wissenschaftsrat (1999). Stellungnahme zur außeruniversitären Materialwissenschaft. Köln.

Wissenschaftsrat (1994). Stellungnahme zur Umweltforschung in Deutschland (zwei Bände). Köln.

Wissenschaftsrat (1999). Leitfaden für die Bewertung von Einrichtungen der Wissenschaftsgemeinschaft Gottfried Wilhelm Leibnitz (WGL) (Blaue Liste). Drs. 3927/99. Köln.

Wissenschaftsrat (1999). Fragebogen für die Bewertung der Forschungseinrichtungen und Museen der Blauen Liste. Drs. 3926/99. Köln.

Wissenschaftsrat (1997). Fragebogen für die Bewertung der Serviceeinrichtungen der Blauen Liste. Drs. 2889/97. Köln.

Hochschulevaluation Deutschland

Barz, Andreas, Carstensen, Doris & Reissert, Reiner (1997). Lehr- und Evaluationsberichte als Instrumente zur Qualitätsförderung. Bestandsaufnahme der aktuellen Praxis. Gütersloh: Centrum für Hochschulentwicklung (Arbeitspapiere, Nr. 13).

Eine Vielfalt von Instrumenten mit den Zielen der Qualitätssförderung in den Hochschulen, der Rechenschaftslegung, der Leistungstransparenz, des Leistungsvergleichs und der Mittelverteilung stehen kaum aufeinander abgestimmt, mit geteilter Akzeptanz und für eine eingeschränkte Öffentlichkeit nebeneinander.
Über die Umsetzung und Nutzung der gesetzlich verordneten Lehrberichte besteht derzeit eine ausgeprägte Ratlosigkeit. Dies liegt u.a. darin begründet, daß sie eher Ausdruck politischer Interessen als eines Konzepts der Qualitätsförderung sind. Es ist unklar, welche Ziele mit den Lehrberichten verfolgt werden sollen, an welche Zielgruppen sie sich richten und was mit den Ergebnissen der Berichte geschehen soll.
Qualitätsförderung nach dem Modell der internen und externen Evaluation von Studium und Lehre bindet Ressourcen in beträchtlichem Ausmaß. Beim derzeitigen Stand der Entwicklung fehlen Follow-up-Untersuchungen, so daß Aussagen über die Nachhaltigkeit der Effekte noch nicht empirisch belegt sind. Kurz- oder mittelfristige Veränderungen sind erzielbar. Zu klären ist allerdings, ob Kosten und Nutzen in einem angemessenen Verhältnis stehen.
Die autonome, wissenschaftliche, profilierte, wettbewerbliche und wirtschaftliche Hochschule bedarf der Konsensbildung über die Ziele der Berichterstattung, die Ebenen der Berichterstattung, der Adressaten, der Inhalte und der damit verbundenen Konsequenzen.
Ein differenziertes und profiliertes Hochschulsystem benötigt Daten über Studierende, Personal, Studien- und Prüfunsgverlauf, Ressourcen, Studienziele und Studienprogramme. Diese dienen der Rechenschaftslegung, der Steuerung, der Planung und der Außendarstellung.
Lehrberichte sind durch den Gesetzgeber verbindlich vorgegeben. Da die Zielsetzung vielfach noch unklar ist, haben die Hochschulen den Freiraum, sie für ihre eigenen Interessen zu nutzen. Lehrberichte bieten den Fächern die Chance, sich ein intern nutzbares Instrument der Organisationsentwicklung zu schaffen. Lehrberichte können eine verbindliche, kontinuierliche Evaluationsform und die Grundlage für die Arbeit in Qualitätszirkeln sein.
Erfolgt die Umsetzung der gesetzlich vorgeschriebenen Lehrberichterstattung in dieser Form, könnten zukünftig Evaluationen nach dem Modell der internen und externen Evaluation gezielt situativ eingesetzt werden als Instrument der Qualitätsüberprüfung etwa bei der Umstrukturierung von Fächern und Hochschulen, bei der Neugestaltung von Studiengängen.
Die fundierte Datenbasis schafft in den Hochschulen durch Evaluation und Berichterstattung die Grundlage für eine gezielte Information der Öffentlichkeit, das heißt Studienplatzbewerber, Hochschulwechsler, aber auch Eltern, Studienberater und Abnehmer werden über die Profile von Studienfächern und das Leistungsspektrum einer Hochschule informiert.

Fischer-Bluhm, Karin (1995). Evaluation im Verbund Norddeutscher Hochschulen. In: Wissenschaftsmanagement 4/95, S. 175-179.

Frackmann, Edgar (1997). Leistungsindikatoren – Das Ende der Debatte. In: Herbert Altrichter, Michael Schratz & Hans Pechar (Hrsg.), Hochschulen auf dem Prüfstand (197-221). Innsbruck: Studien-Verlag.

Der Autor zeigt, dass Leistungsindikatoren nicht mehr bloß Gegenstand theoretischer Diskussionen sind, sondern dass sie sich im Hochschulbereich bereits praktisch etabliert haben. Der Beitrag beschreibt zwei Tendenzen: Zum einen sind diese Indikatoren Bestandteil externer, formelgebundener Allokationsentscheidungen auf verschiedenen Ebenen, zum anderen dienen sie als Information zur Prozessgestaltung im Rahmen des selbstbestimmten Qualitätsmanagements von Universitäten.

Hochschulrektorenkonferenz und Hochschul-Informations-System GmbH (o. J.). EvaNet: Evaluations-Netzwerk zur Evaluation und Qualitätssicherung. http://evanet.his.de/evanet/

Hochschulrektorenkonferenz (1999). Viel Lärm um nichts? Evaluation von Studium und Lehre und ihre Folgen. Tagung an der Universität Rostock vom 6. bis 8. September 1998. Bonn: HRK (Projekt Qualitätssicherung, 4/1999).

Hochschulrektorenkonferenz (1999). Ein Schritt in die Zukunft. Qualitätssicherung im Hochschulbereich. Berliner Bildungsdialoge. Hochschulrektorenkonferenz und Veranstaltungsforum der Verlagsgruppe Georg von Holtzbrinck. Berlin, 26. Oktober 1998. Bonn: HRK (Projekt Qualitätssicherung, 3/1999).

Hochschulrektorenkonferenz (1999). Qualität an Hochschulen. Fachtagung der Universität Kaiserslautern und der Hochschulrektorenkonferenz. Kaiserslautern, 28./29. September 1998. Bonn: HRK (Projekt Qualitätssicherung, 1/1999.

Hochschulrektorenkonferenz (1998). Evaluation und Qualitätssicherung an den Hochschulen in Deutschland - Stand und Perspektiven. Nationales Expertenseminar der Hochschulrektorenkonferenz. Bonn, 29. Mai 1998. Bonn: HRK (Projekt Qualitätssicherung, 6/1998).

Kieser, Alfred (1998). Going Dutch - Was lehren niederländische Erfahrungen mit der Evaluation universitärer Forschung? DBW, 58(2), 208-224.

Küpper, Hans-Ulrich (1996). Struktur, Aufgaben und Systeme des Hochschul-Controlling. Beiträge zur Hochschulforschung, 3-1996, 147-179.

Daten über Input- und Outputgrößen sind eine wertvolle Basis für die Analyse der Geschäftsprozesse. Damit kann durch ihre Ermittlung der Diskurs über die Gestaltung dieser Prozesse und die sie bestimmenden Entscheidungen auf Hochschul- und Fakultätsebene angestoßen werden sowie unterstützt werden. Für den Vergleich und die Beurteilung der Prozesse wird es notwendig, Kennzahlen z. B. aus dem Verhältnis zwischen Output- und Inputgrößen (als eine Art von Produktivitätskennzahlen) zu bilden, die als Indikatoren für deren Effektivität, Qualität oder Effizienz deutbar sind. Da sich die Forschungs-, Lehr- und Serviceprozesse nicht in ökonomischen Größen bewerten lassen, ist ein intensiver Diskussionsprozeß über den Gehalt verschiedener Kennzahlen zweckmäßig, in dem sich die als relevant erachteten Zielgrößen herausschälen. Durch die Schaffung von Transparenz und die Auslösung von Diskussionsprozessen wird darüber hinaus das Wertesystem und die Kultur der Hochschule beeinflußt.

Montada, Leo, Krampen, Günter & Burkard, Patrick (1999). Persönliche und soziale Orientierungslagen von Hochschullehrern/innen der Psychologie zu Evaluationskriterien über eigene berufliche Leistungen. Psychologische Rundschau, 50(2), 69-89.

Müller-Böling, Detlef (1997). Evaluationen zur Rechenschaftslegung oder Qualitätsverbesserung? Eine Bestandsaufnahme der Evaluation an deutschen Hochschulen.

In: Herbert Altrichter, Michael Schratz & Hans Pechar (Hrsg.), Hochschulen auf dem Prüfstand (88-107). Innsbruck: Studien-Verlag.

Evaluation ist zu *dem* Modewort der hochschulpolitischen Entwicklung geworden. Vor einigen Jahren in Deutschland im Zusammenhang mit Hochschulen noch völlig ungebräuchlich, scheint sich damit nunmehr für viele die Lösung aller Probleme des tertiären Bildungsbereichs zu verbinden. Selbstverständlich ist die Begutachtung und die Bewertung in Hochschulen kein unbekannter Vorgang, dennoch verbirgt sich hinter diesem Sammelbegriff nunmehr eine Vielzahl von unterschiedlichen neuen Ansätzen.

Ott, Robert (1999). Darstellung und Beurteilung von Hochschul-Rankings in Deutschland. Beiträge zur Hochschulforschung, 4-1999, 309-322.

Das Schlagwort „Wettbewerb“ taucht immer häufiger in der derzeitigen Diskussion über eine Reform des deutschen Hochschulwesens auf. Die Hochschulen werden dabei nicht mehr als einheitlich angesehen, sondern es werden die Unterschiede in der Forschung und Lehre durchaus wahrgenommen. Da es an einem einheitlichen Bewertungs- und Vergleichsmaßstab mangelt, haben sich seit 1989 verschiedene Magazine dieses Problems angenommen und eine Vielzahl von Ranglisten, die sogenannten Rankings, erstellt. Dieser Beitrag hat zum Ziel, die bisher veröffentlichten Rankings in Deutschland darzustellen, zu systematisieren und auf ihre Aussagekraft und Vergleichbarkeit hin zu untersuchen. Trotz des dabei gewonnenen Ergebnisses, dass die Ranglisten wegen ihrer unterschiedlichen Methodik miteinander nicht vergleichbar und zudem wenig aussagekräftig sind, wurden die Rankings bezüglich ihrer Platzierungen untersucht und festgestellt, dass eine gewisse Konstanz innerhalb der jeweiligen Gruppe von Befragten vorherrscht.

Unter Bezug auf die Bestrebungen und Erfordernisse nach einer Evaluation - der Qualitätssicherung und dem Qualitätsmanagement - der Arbeit von Hochschullehrern(innen) und Universitätsinstituten wurde eine Expertenbefragung durchgeführt, um die Frage zu klären, ob für dabei potentiell verwendbare Evaluationskriterien in der Gruppe der an deutschen Psychologischen Instituten mit Hauptfachausbildung tätigen Hochschullehrern(innen) ein kollegialer Konsens existiert. Präsentiert werden Befunde, die in einer Stichprobe von 265 Experten in eigener Sache gewonnen wurden und sich auf die persönlichen Bewertungen von 117 zur Beurteilung vorgegebenen möglichen Evaluationskriterien, die darauf bezogenen sozialen Orientierungslagen und die Konsensfähigkeit der einzelnen Kriterien sowie die Dimensionalität der auf die Kriterien bezogenen Wichtigkeitseinschätzungen beziehen.

Rat für Wissenschaft und Forschung (1999). Wirtschaftswissenschaften an den Bayerischen Universitäten: Evaluierungsbericht. München: Bayerisches Staatsministerium für Unterricht, Kultus, Wissenschaft und Kunst.

Der Rat für Wissenschaft und Forschung des Bayerischen Staatsministers für Unterricht, Kultus, Wissenschaft und Kunst ist vom Staatsminister gebeten worden, eine Begutachtung des Ist-Zustandes der Wirtschaftswissenschaften an den bayerischen Landesuniversitäten durchzuführen und Empfehlungen zu ihrer künftigen Ausgestaltung auszuarbeiten. Der Rat hat diesen Auftrag dem Ausschluss Wirtschaftswissenschaften übertragen (Vorsitz: Frau Prof. Dr. Renate Schubert, ETH Zürich). Der Auftrag des Ausschusses verlangte in einem ersten Schritt die Erhebung des gegenwärtigen Standes der Wirtschaftswissenschaften an den bayerischen Landesuniversitäten hinsichtlich der Lehre und Forschung sowie der dort jeweils vertretenen Schwerpunkte. Die so gewonnenen Erkenntnisse waren danach zu bewerten, (1) in welcher Differenzierung sich die Wirtschaftswissenschaften an den einzelnen Universitäten darstellen, (2) wie fundiert, effizient und zeitgerecht das Lehrangebot ausgestaltet ist, (3) wie die Wirtschaftswissenschaften am aktuellen Forschungsgeschehen teilnehmen und dieses mitgestalten und (4) wie offen ihre Strukturen für die Aufnahme neuer Entwicklungen ist. In diesem Zusammenhang war auch die Angemessenheit und Erforderlichkeit der jeweils vorhandenen personellen und sonstigen Ausstattung zu überprüfen.

Reissert, Reiner & Carstensen, Doris (1998). Praxis der internen und externen Evaluation. Handbuch zum Verfahren. Hannover: HIS Kurzinformation.

Schlink, Bernhard (1999). Evaluierte Freiheit? Zu den Bemühungen um eine Verbesserung der wissenschaftlichen Lehre. Berlin: Humboldt Universität zu Berlin (Reihe Öffentliche Vorlesungen, Heft 100).

Der Autor fragt, was Artikel 5, Absatz 3 des Grundgesetzes an Bewertungen wissenschaftlicher Forschung und Lehre zuläßt und ob dies für eine sich reformierende Universität eine Sperre und Bürde ist oder den richtigen Weg weist.

Schoder, Thomas (1999). Budgetierung als Koordinations- und Steuerungsinstrument des Controlling in Hochschulen. München: Bayerisches Staatsinstitut für Hochschulforschung und Hochschulplanung (Monographien: Neue Folge, Band 54).

Ausgehend von verschiedenen Ansatzpunkten wächst die Zahl der Untersuchungen, die sich mit der Mittelverteilung in Hochschulen beschäftigen, stetig. Die vorliegende Arbeit untersucht die Budgetierung in Hochschulen aus dem spezifischen Blickwinkel der Principal-Agent-Theorie heraus. Damit wird der Fokus auf Interessenkonflikte und Informationsasymmetrien gelenkt, die im Zusammenhang mit der Budgetierung in Hochschulen von Bedeutung sind. Zwei mögliche Wege einen Interessenausgleich herbeizuführen sind die Bereitstellung von Ressourcen für künftig zu erbringende Leistungen und die Zuweisung von Mitteln für erbrachte Leistungen und Erfolge. Es ist zu erwarten, dass die zukünftige Bedeutung und Entwicklung der Budgetierung in Hochschulen vor allem durch Zielvorgaben und Zielvereinbarungen beeinflusst werden wird. Voraussetzung hierfür ist eine Verständigung auf relevante Ziele innerhalb der Hochschule und deren Operationalisierung in Kennzahlen- bzw. Zielsystemen.

Sieber, W. (Hrsg.). (1996). Evaluation an Fachhochschulen – Chancen und Risiken. Neuenrade: Hochschullehrerbund NRW (Schriften des Hochschullehrerbundes, Landesverband Nordrhein-Westfalen, Band 2).

Teichler, Ulrich & Schomburg, H. (1997). Evaluation von Hochschulen auf der Basis von Absolventenstudien. In: Herbert Altrichter, Michael Schratz & Hans Pechar (Hrsg.), Hochschulen auf dem Prüfstand (235-260). Innsbruck: Studien-Verlag.

Die Autoren beleuchten zunächst unterschiedliche Forschungsrichtungen zur Messung der Qualität von Hochschulen und arbeiten die Stärken und Schwächen der Forschung über Rangstufungen heraus. Den Hauptteil des Beitrages machen Absolventenstudien als Instrument zur Analyse der Vielfalt von Hochschulen und deren Wirkungen aus. Sie untersuchen systematisch Zusammenhänge zwischen den Studienbedingungen und –angeboten einerseits und Berufsweg und –tätigkeit der Absolventen andererseits. Daraus leiten sie Überlegungen für ein Indikatorensystem zur Messung der Leistung von Hochschulen anhand der Erforschung des Übergangs vom Studium zur Berufstätigkeit und der Berufstätigkeit in den ersten Jahren nach Studienabschluss ab.

Voegelin, Ludwig (1997). Evaluation im Verbund norddeutscher Hochschulen. Wirksamkeitsanalyse. In: HIS A 12/97: Wirksamkeit der interen und externen Evaluation von Studium und Lehre, S.2-8.

Weichselbaumer, Jürgen Stefan (1999). Hochschulrechnungswesen im Wandel. Entwicklungen, Bestandsaufnahmen, Perspektiven. Beiträge zur Hochschulforschung, 4-1999, 279-293.

Die Hochschulreform in Deutschland schreitet unaufhaltsam voran. Kernprinzipien wie Wettbewerb, Autonomie und Leistungsorientierung bedingen veränderte Informationsanforderungen an die Hochschulen, denen das Hochschulrechnungswesen Rechnung tragen muss. Im vorliegenden Beitrag werden aktuelle Entwicklungen im Umfeld der Hochschulen mit ihren Konsequenzen für das Hochschulrechnungswesen skizziert. Ergebnisse einer Befragung der deutschen Universitätskanzler und die Greifswalder Grundsätze zum

Hochschulrechnungswesen zeichnen ein Bild von der bestehenden Situation an den deutschen Universitäten. Die Ausführungen münden in einem Szenario für die weitere Hochschulentwicklung, das die knappe Finanzsituation an den Hochschulen zum Ausgangspunkt hat.

Wex, Peter (1995). Die Mittelverteilung nach Leistungs- und Belastungskriterien. Ein Beitrag zum Leistungswettbewerb in der Hochschule. Wissenschaftsmanagement, 1(4), 168-174.

Wissenschaftsrat (1996). Thesen zur Forschung in den Hochschulen (These 7: Ressourcenvergabe und Evaluation, S. 61-66). Köln. (Drs. 2765/96 pi)

Der Wettbewerb um Drittmittel ist der wichtigste Weg für eine leistungsbezogene Vergabe von Forschungsressourcen. Das Drittmittelvolumen muß daher gesteigert werden. Zusätzlich sollten die Hochschulen eigene Instrumente für eine qualitätsabhängige Ressourcenverteilung schaffen. Angemessene Formen interner und externer Evaluation sind nützlich, um unterschiedliche Ressourcenzuweisungen zu begründen; die Erfüllung von Evaluationskriterien darf aber nicht zum Selbstzweck werden.

Wissenschaftsrat (1996). Empfehlungen zur Stärkung der Lehre in den Hochschulen durch Evaluation. Köln.

Witte, Frank (1999). Wirtschaftlichkeit in Hochschulen. Ein Finanzmittelverteilungsmodell für Hochschulen auf der Grundlage der Lehr- und Forschungsleistung. Aachen: Shaker Verlag.

Der Autor entwickelt ein Konzept für ein Verteilungsmodell, welches sich auf die theoretischen Grundlagen eines mehrstufigen Principal-Agent-Modells stützt. Dabei zeigt sich, dass leistungsorientierte Finanzmittelzuweisungen Anreize zur Ausrichtung an den generellen gesellschaftspolitischen Zielen bieten. Die aufgaben- und belastungsorientierten Zuweisungen sowie die leistungsinduzierten Zuweisungen basieren auf Ex-post-Informationen, während sich die Innovationsförderung auf Ex-ante-Informationen stützt. Die aufgaben- und belastungsorientierte Zuweisung dient der Finanzierung der Grundausstattung, die leistungsinduzierte Zuweisung soll Anreize zur Leistungssteigerung bieten. Grundlage der beiden letztgenannten Zuweisungen sind Indikatoren, wobei sich die Leistungsindikatoren auf die Urteile der Scientific Community stützen.

Wolff, Brigitta (1995). Organisation durch Verträge. Wiesbaden: Deutscher Universitäts-Verlag.

6. Niederlande

Forschungsevaluation

Beemt, F. C. H. D. van den & Pair, C. le (1991). Grading the grain: consistent evaluation of research proposals. Research Evaluation, 1(1), 3-10.

The article describes the selection procedure for research proposals which was introduced by the Technology Foundation of the Netherlands in 1981. The procedure involves obtaining peer reviews, then the responses of the principal investigator are added to the review document. Twenty such proposal documents are then judged by a non-peer jury.

Caulil, G. F. van Mombers, C. A. M. & van den Beemt, F. C. H. D. (1996). Quantifying the utilization of research: the difficulties and two models to evaluate the utilization of research results. Scientometrics, 37(3), 433-444.

To evaluate utilization in a broad sense one model describes the degree of utilization with three aspects: the involvement of the user, the availability of a transferable research product, and the commercial benefits resulting from the reseach results. In the other method the utilization of the research results is described first, and subsequently the utilization is quantified by a jury, who group the different projects in five classes, based on a Guttman scale.

Eiffinger, Marcel A. M. (1997). Evaluation of Scientific Research in The Netherlands. In OECD (Ed.), The Evaluation of Scientific Research: Selected Experiences (27-46). Paris: OECD/GD(97)194.

Glänzel, Wolfgang, Rinia, E. J. & Brocken, M. G. M. (1995). A bibliometric study of highly cited European physics papers in the 80s. Research Evaluation, 5(2), 113-122.

Highly cited publications are determined using citations from The SCI database and source documents from a non-SCI database. Instead of the journal-impact factor, which in this case is questionable as a reference standard, the average citation rate of a subfield was used for the analysis of European physics papers published in the period 1980-87 and covered by the Physics Briefs database. The added value of indicators based on highly cited papers to other indicators such as the citation mean is shown. The paper concludes with an investigation of the role of international collaboration among highly cited papers.

Glänzel, Wolfgang, Schubert, A., Braun, T., Rinia, E. J. & Brocken, M. G. M. (1994). Bibliometric indicators of European physics in the 80s. Utrecht: Foundation for Fundamental Research on Matter, Report FOM-94.1483.

Heeringen, A. van (1999). The role of advisory bodies in evaluation. European Commission & Austrian Advisory Board for Universities (Eds.), Science and the Academic System in Transition: An International Expert Meeting on Evaluation (59-61). Budapest: Akadémiai Kiadó.

Jansz, C. N. M. & Le Pair, C. (1992). Bibliometric invisibility of technological advances. In: Peter Weingart et al. (Eds.), Representations of Science and Technology (315-326). Proceedings of the International Conference on Science and Technology Indicators. Leiden: DSWO Press.

Kaa, D. J. van de (1994). Picking the winners by concensus: Grant-giving practice in the Netherlands. Higher Education, 28, 59-83.

Korevaar, J. C. & Moed, H. F. (1996). Validation of bibliometric indicators in the field of mathematics. Scientometrics, 37, 117-130.

A survey was conducted to answer the following question: to which extent to citation-scores mirror to the opinions of experts concerning the quality of a paper or a journal? In conclusion, the experts' views on top publications or top journals correspond very well to bibliometric indicators based on citation counts.

Leeuwen, Th. N. van, Rinia, E. J. & Raan, A. F. J. van (1996). Bibliometric Profiles of Academic Physics Research in the Netherlands. Research Report to the Foundation for Fundamental Research on Matter (FOM). Report CWTS 96-09.

The report presents the results of a study on the research performance of Dutch physics carried out in 220 research programmes at universities and research institutes in the Netherlands. The aim of the study is to analyse the production, impact and visibility of Dutch physics research in an international perspective. Based on the complete publication oeuvre in the period 1985-1994 of 722 senior scientists participating in these programmes in 1995, an assessment is made of the set of programmes as a whole, of programmes at the institutional level, at the funding level and at the field level. The publication output, impact and scientific collaboration was studied using bibliometric techniques.

Le Pair, C. (1997). Formal Evaluation Methods: Their Utility and Limitations. In Mark S. Frankel & Jane Cave (Eds.), Evaluating Science and Scientists: An East-West Dialogue on Research Evaluation in Post-Communist Europe (170-182). Budapest: Central European University Press.

In the case of technology, as distinguished from basic science, the author argues that reliance on quantitative indicators is an insufficient means of evaluation and can be misleading. To support his argument, le Pair identifies a 'Citation Gap' in various fields where 'innovative technology' leading to practical uses is 'bibliometrically nearly invisible'. Le Pair concludes his paper with a description of how the technology branch of the Netherlands Research Council evaluates proposals and awards grants. After proposals are reviewed by experts in the relevant fields, they are then scrutinized by an ad hoc lay jury consisting of 11 people from universities, other government institutions, and private companies, which is asked to score each proposal for scientific merit and for the likelihood of useful application. In this system, which has been in place since 1980, jury judgments regarding the likely utility of proposed research have proved to be highly predictive.

Leydesdorff, Loet (1999). The Challenge of Scientometrics: The development, measurement, and self-organization of scientific communication. Leiden: DSWO Press.

The various dimensions of the problem of studying the sciences empirically are clarified in the book in a methodological analysis of theoretical traditions, including the sociology of scientific knowledge and neo-conventionalism in the philosophy of science.

Leydesdorff, Loet & Gauthier, Élaine (1996). The evaluation of national performance in selected priority areas using scientometric methods. Research Policy, 25, 431-450.

The paper is based on a comparative study of strategic research programs in Canada and the Netherlands.

Leydesdorff, Loet & Schaar, P. van der (1987). The use of scientometric methods for evaluating national research programs. Science and Technology Studies, 5(1), 22-32.

Moed, H. F. (1996). Differences in the construction of SCI based bibliometric indicators among various producers: a first overview. Scientometrics, 35(2), 177-191.

Moed, H. F. & Hesselink, F. Th. (1996). The publication output and impact of academic chemistry research in the Netherlands during the 1980's: bibliometric analyses and policy implications. Research Policy, 25, 819-836.

Moed, H. F., Bruin, R. E. de, Leeuwen, Th. N. (1995). New bibliometric tools for the assessment of national research performance: Database description overview of indicators and first applications. Scientometrics, 33, 381-422.

Nederhof, A. J. & Wijk, van E. (1997). Mapping the social and behavioral sciences worldwide: use of maps in portfolio analysis of national research efforts. Scientometrics, 40(2), 237-276.

Nederhof, A. J. (1996). A bibliometric assessment of research council grants in linguistics. Research Evaluation, 6(1), 2-12.

The research performance of researchers supported by its grants ('grantees') was compared with that of other researchers ('non-grantees'). Although publications in ISI source journals were cited about twice as often as other publication types, the latter were produced more often. Grantees were more likely to be highly productive and produced significantly more highly cited articles than non-grantees. Although both grantees and non-grantees were cited well above the international level, articles by grantees showed a faster diffusion, and had a higher impact in both core and peripherial journals.

Nederhof, A. J. & Straathof, A. (1994, revised version). Productivity and Impact of Economic Sciences in the Netherlands. Research Report to the Netherlands Organization for Scientific Research (NWO), The Hague. Report CWTS-93-02

Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Ed.). (1997). The Future of the Peer Review System. The Hague: NWO.

The booklet contains three papers on the peer review system by Richard Brook (UK), Arie Rip (The Netherlands) and Fiona Wood (Australia) as it related to the award of research grants.

Noyons, Ed C. M. (1999). Bibliometric Mapping as a Science Policy and Research Management Tool. Leiden: DSWO Press.

Bibliometric maps of science are landscapes of scientific fields created by quantitative analysis of bibliometric data. As a policy supportive tool bibliometric maps can be applied to overview the structure of a research field and to monitor its evolution. A methodology and procedure is proposed to allow both expert (trustworthiness) and user (utility) to validate the maps. Furthermore, a procedure is outlined to monitor actor’s activities in view of the international developments in a research field. The proposed methodology opens new doors for evaluative bibliometrics and is prepared for the advent of electronic publishing in science.

Noyons, Ed C. M. & Raan, A. F. J. (1999). Combining Mapping and Citation Analysis for Evaluative Bibliometric Purposes: A Bibliometric Study. Journal of the American Society for Information Science, 50(2), 115-131.

The aim of the article is to demonstrate how the results both of a structural analysis, and of a research performance assessment of a research field, can be enriched by combining elements of both into one integrated analysis. In addition, a procedure is discussed to select and analyse candidate benchmark institutes to assess the position of a particular research institute, in terms of both its cognitive orientation and its scientific production and impact at the international research front.

Noyons, Ed C. M. & Raan, A. F. J. (1996). Actor analysis in neural network research: the position of Germany. Research Evaluation, 6(2), 133-142.

The results of a bibliometric study of neural network research are presented. The evaluative study includes bibliometric mapping and actor analysis of main players in the field on a macro level (countries, in paricular Germany), and, on a lower level, of the main players in Germany. We found that Germany is among the leading countries in the field. The study is also a blueprint for evaluative bibliometric studies of emerging or strongly developing science and technology fields. The monitoring of major developments in the fields, and a detailed actor analysis were integrated into one study.

Netherlands Organisation for Scientific Research (1998). Guidelines for the evaluation of NWO Institutes. Den Haag: NWO.

Netherlands Organisation for Scientific Research (1996). Report of the NWO Evaluation Committee. Den Haag: NWO.

The Netherlands Organisation for Scientific Research (NWO) Act prescribes a periodical evaluation process for the organisation as a whole and its constituent parts. Thus, in 1995, the Minister of Education, Cultures and Sciences and the Governing Board of NWO decided that the time had come to evaluate NWO's performance during the first seven years of existence. An international evaluation committee was appointed and provided with a list of ten questions addressing various aspects of NWO's activities.

Raan, Anthony F. J. van (1999). Evaluation of Performance and Trends in Basic and Applied Research by Advanced Bibliometric Methods: A Science Policy Instrument for Nations with an Economy in Transition. In Vaclav Paces, Ladislav Pivec & Albert H. Teich (eds.), Science Evaluatiion and Its Management (227-245). Ohmsha: IOS Press, NATO Science Series: Science & Technolgy Policy, vol. 28.

The paper presents an overview of the potentials and limitations of bibliometric methods for the assessment of strengths and weaknesses in research performance, and for monitoring scientific developments. Two modalities of application are addressed: assessment of a large research institute (institutional level), and assessment of an entire scientific discipline, nation-wide, by university and by field (national level). The findings of the bibliometric studies provide insight into the international position of actors at the research front in terms of influence and specialisations, as well as into patterns of scientific communication and processes of knowledge dissemination.

Raan, Anthony F. J. van (1998). Assessment of social sciences: the use of advanced bibliometric methods as a necessary complement of peer review. Research Evaluation, 7(1), 2-6.

In the paper it is argued that bibliometric performance indicators allow substantial improvement of peer-review based evaluation in the social sciences. Advanced bibliometric indicators provide up-to-date, detailed, 'objective' and structured information on the performance (particularly 'impact') of a research group. They prevent the peer-review process from becoming too soft or too 'uninterested'. They also provide information not readily made available by peer review.

Raan, Anthony F. J. van (1997). Scientometrics: state-of-the-art. Scientometrics, 38(1), 205-218.

Raan, Anthony F. J. van (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. Scientometrics, 36, 397-420.

Raan, Anthony F. J. van (1993). Advanced bibliometric methods to assess research performance and scientific development: basic principles and recent practical applications. Research Evaluation, 3(3), 151-166.

A broad overview is presented of quantitative methods, based on bibliometric data. Specific indicators for important aspects of scientific research can be constructed, the main types being characteristics of scientific output; characteristics of scientific impact; and maps of science.

Bibliometric indicators are particularly useful in the evaluation of natural sciences and medical research, but they also have increasing potential in the applied and engineering sciences, the humanities, and in the social sciences. Bibliometric mapping enables us to visualise scientific and technological developments, and the actors who play important roles in the different (sub)fields. The performance measurement and the mapping methods together provide a powerful analytic tool to assess research activities and to monitor science and technology.

Rinia, E. J. (2000). Scientometric studies and their role in research policy of two research councils in the Netherlands. Scientometrics, 47(2).

In the past 30 years a variety of scientometric analyses have provided input data for research policy objectives of research institutions in the Netherlands. In the article the author discuss several pioneering studies performend on behalf of the research councils for physics (FOM) and technical sciences (STW), which have played an important role in the early development of scientometrics in the Netherlands. The motives for these studies, the results and the influence on research policy are discussed. Relations to present themes in scientometric investigations are drawn.

Rinia, E. J., Leeuwen, Th. N. van, Vuren, H. G. van & Raan, A. F. J. (1998). Comparative analysis of a set of bibliometric indicators and central peer review criteria. Evaluation of condensed matter physics in the Netherlands. Research Policy, 27, 95-107.

The paper reports first results of a study on the correlation between bibliometric indicators and the outcomes of peer judgements made by expert committees of physics in the Netherlands. As a first step to study these outcomes in more detail, we focus on the results of an evaluation of 56 research programmes in condensed matter physics in the Netherlands, a subfield which accounts for roughly one third of the total of Dutch physics. This set of research programmes is represented by a volume of more than 5,000 publications and nearly 50,000 citations. The study shows varying correlations between different bibliometric indicators and the outcomes of a peer review procedure.

Rip, Arie (1999). Societal Challenges for Evaluation. In: Susanne Bührer & Stefan Kuhlmann (Eds.), Evaluation of Science and Technology in the new Europe. Proceedings of an International Conference on 7 and 8 June 1999, Berlin (127-131). Bonn / Bruxelles: Federal Ministry of Education and Research / European Commission.

Rip, Arie (1997). Higher Forms of Nonsense. In: Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Ed.), The Future of the Peer Review System (27-51). The Hague: NWO.

The paper describes five scenarios for the future of peer review in relation to funding agencies.

Rip, Arie (1997). Qualitative conditions of scientometrics: the new challenges. Scientometrics, 38(1), 7-26.

Rip, Arie & Meulen, Barend J. R. van der (1995). The patchwork of the Dutch evaluation system. Research Evaluation, 5(1), 45-53.

In the Netherlands, an evaluation culture has evolved. Systematic evaluation, however, occurs only for strategic R&D programmes and innovation-oriented programmes, and as part of a quality-assurance system of academic research. Informal assessments and bottom-up evaluation activities are dominant in the Dutch approach. Science policy agencies have been interested in strategic changes in the research system, rather than in evaluation. They have also stimulated the development of an infrastructure for evaluation, rather than concentrating on assessing their own activities.

Rip, Arie (1990). Implementation and Evaluation of Science & Technology Priorities and Programs. In Susan Cozzens, Peter Healey, Arie Rip & John Ziman (eds.), The Research System in Transition (263-280). Dordrecht: Kluwer.

Royal Netherlands Academy of Arts and Sciences (1998). Discipline Report on (Bio)Mecial and Health Sciences Research in the Netherlands. Amsterdam: KNAW.
http://www.knaw.nl/cg/homedag.htm
http://www.knaw.nl/09public/0903a.htm
http://www.knaw.nl/09public/0903b.htm

Royal Netherlands Academy of Arts and Sciences (1997). Externe evaluatie Koninklijke Nederlandse Akademie van Wetenschappen. Amsterdam: KNAW.
http://www.knaw.nl/09public/0903c.htm

Royal Netherlands Academy of Arts and Sciences (1997). Evaluatie van het onderzoek op het gebied van de geesteswetenschappen. Amsterdam: KNAW.
http://www.knaw.nl/09public/0903b.htm

Royal Netherlands Academy of Arts and Sciences (1997). A sense of direction - The future of legal research in the Netherlands. Amsterdam: KNAW. ISBN 90-9010326-0

Royal Netherlands Academy of Arts and Sciences (1995). Peer review inzake de systematische biologie in Nederland. Amsterdam: KNAW.
http://www.knaw.nl/09public/0903b.htm

Spaapen, Jack & Sylvain, Christian (1993). Assessing the value of research for society. Research Evaluation, 3(2), 117-126.

Many of the choices that have to be made in the research process are stimulated or frustrated by the contingencies in the societal context, yet it remains unclear as to how precisely to make those choices. The issue of how to assess research in that context is even more ambiguous. The paper reports on a study commissioned by the government on the Netherlands to find a more comprehensive form of research evaluation, in which the dynamic interaction between research and society is represented. It criticizes traditional approaches of evaluation, and presents a framework in which different indicators are integrated that refer to research output, societal demand and mediated interactions.

Steen, Jan van (1996). A government's S&T policy strategy to cope with stagnating R&D budgets. Research Evaluation, 6(2), 77-82.

Some new elements of the Dutch government's S&T policy in the 90s are described, which focus on tackling stagnating R&D budgets. First, a general view of the Dutch R&D situation is given; second, attention is paid to some analytical work on government R&D funding in the context of budgetary constraints; and third some new elements - priority setting, the interaction between science and society, and the current government strategy for fostering R&D and innovation - are discussed.

Steen, Jan van (1995). S&T indicators in science policy: how can they matter? Research Evaluation, 5(2), 161-166.

S&T indicators can have a number of objectives in science and technology policy, of which monitoring and evaluation are the most important. The main objective of the current indicators is to monitor the S&T system in a general way, but it will be necessary to optimise the potential of S&T indicators to generate information that can be used in more specific discussions on S&T policy.

Steen, Jan van & Eijffinger, Marcel (1998). Evaluation practices of scientific research in the Netherlands. Research Evaluation, 7(2), 113-122.

The article takes stock of Dutch evaluation practices for publicly funded basic and strategic research in three different contexts: institutional strategy formulation; allocative decision-making; and research and science policy strategy. The different evaluation practices are dealt with in detail, by describing their technical set-up. In conclusion, the main challenge for science policy over the

next few years is discussed: how can the different evaluation practices be bridged, including the integration of the societal point of view within evaluation practice.

Van der Meulen, B. R. (1997). The use of S&T indicators in science policy: Dutch experiences and theoretical perspectives from policy analysis. Scientometrics, 38(1), 87-101.

The relationship between bibliometrics and science policy remains underdeveloped. Relevance of new methods to produce indicators is easily claimed, but often without real insight in the policy processes. Drawing on experiences with the use of S&T indicators in science policy in the Netherlands and on principal-agent theory, the author develops an analytical perspective which enables to assess the role of S&T indicators in science policy. It is argued that the use of S&T indicators can only be understood well if one takes the socio-political context with its specific dynamics and rationalities into account.

Van der Meulen, B. R. & Rip, A. (1998). Mediation in the Dutch science system. Research Policy, 27, 757-769.

Van der Meulen, B. R. & Rip, A. (1997). Social quality of research between responsibility and management: An inventory of evaluation practices (mimeo). Enschede: University of Twente.

Social quality is defined as the extent to which research (can be expected to) contribute to desired social developments, to knowledge of social importance and to aims that are deemed important in society or in specific social sectors.

Verkleij, Adrian C. L. (1999). Different approaches to defining research quality. Bulletin, vol. 5, 2-6 (ISSN 10176135).

Standards for assessing research often favour research with international impact. This traditional view on quality of research has come under criticism by those who see no longer fundamental research as the sole and exclusive research mission of a university. Many universities have developed research portfolios which contain mixtures of fundamental research, applied research, technology, development work, etc. However, assessment practices tend to assess the complete portfolio along the criteria of fundamental research. Scholars involved with non-fundamental research activities often come out with lower ratings, because they do not conform to the (publication) standards set by the representatives of the fundamental sciences. Therefore, the author suggest the introduction of a more contextual approach to quality, based on 'fitness for purpose'. This separates the appraisal of purposes (which is highly political) from an assessment of the capacity of a research group to fulfil the promises made in their purposes.

Verkleij, Adrian C. L. (1999). Self-evaluation and External Review. In: Röbbecke, Martina & Simon, Dagmar (Hrsg.). Qualitätsförderung durch Evaluation? Ziele, Aufgaben und Verfahren von Forschungsbewertungen im Wandel (87-99). Berlin: Wissenschaftszentrum Berlin für Sozialforschung (P 99 – 003).

Wouters, P. (1999). The Citation Culture. Ph.D. Thesis. Amsterdam: University of Amsterdam, Department of Science Dynamics.

Wouters, P. (1997). Citation Cycles and Peer Review Cycles. Scientometrics, 38(1), 39-55.

Science is pictured as an information processing cycle. Its quality us maintained in the peer review cycle. The main upshot of the Science Citation Index (SCI) has been the creation of a second-order cycle on top of the primary knowledge production cycle.

Hochschulevaluation Niederlande

Bruggen, Johan C. van, Scheele, Jacob P. & Westerheijden, Don F. (1999). To be continued... Syntheses and trends in follow-up of quality assurance in West European higher education. European Journal for Educational Law and Policy, 2, 155-163.

The authors report on West European trends in follow-up arrangements of national quality assurance procedures. They argue that external quality assurance ought to adopt a broader conception of quality than is now often the case. Next, the general public ought to have a guarantee that external quality assurance is valid (through meta-evaluation) and has consequences (through follow-up arrangements). Follow-up currently is not well developed in many European countries in a formal sense, but is receiving more attention as external quality assurance is maturing. Finally, the authors link national quality assurance to the European dimension; an international network could fulfil important functions in making quality assurance transparent and credible across Europe.

Jongbloed, Ben & Knoop, Han van der (1999). Budgeting at the institutional level: Responding to internal pressures and external opportunities (141-164). In Ben Jongbloed, Peter Maassen & Guy Neave (Eds.), From the Eye of the Storm: Higher Education's Changing Institution. Dordrecht: Kluwer Academic Publishers.

Moed, H. F. (1996). Differences in the construction of SCI based bibliometric indicators among various producers: a first overview. Scientometrics, 35(2), 177-191.

Moed, H. F., Leeuwen, Th. N. van & Visser, M. S. (1999). Trends in publication output and impact of universities in the Netherlands. Research Evaluation, 8(1), 60-67.

Trends in the research output and impact of universities in the Netherlands are examined, as reflected in scientific articles in journals processed for the Science Citation Index. At the level of subfields in the natural, technical and life sciences, there has been hardly any concentration of research activities among Dutch universities during the 80s and 90s. In the 80s and to a lesser extent in the 90s, levelling of univesities' research output in natural and life sciences is a dominent trend. Changes in distribution of students among the universities and the outcomes of evaluation studies conducted in the past probably have affected this trend positively towards uniformity in output. The academic systems in Sweden, Italy, Germany, Spain, Denmark and particularly Great Britain show a stronger concentration of research articles among universities than the Dutch academic system.

Richter, Roland (1997). Qualitätsevaluation an niederländischen Universitäten und Fachhochschulen. In: Herbert Altrichter, Michael Schratz & Hans Pechar (Hrsg.), Hochschulen auf dem Prüfstand (108-122). Innsbruck: Studien-Verlag.

In den Niederlanden legte das Wissenschaftsministerium bereits 1985 den Bericht „Hoger Onderwijs: Autonomie en Kwaliteit“ (HOAK) vor, der eine größere Autonomie der Hochschulen befürwortete und die Entwicklung eines umfassenden und inzwischen durch das Hochschulgesetz von 1993 (WHW) rechtlich abgesicherten Qualitätsmanagements an den Universitäten und Fachhochschulen forderte. Die Verantwortung für die Sicherung der Qualität und die Organisation des Lehrangebots und der Forschung liegt bei den Fakultäten und Hochschulen, die von Anfang an akzeptiert haben, dass Autonomie und Qualitätsevaluation die beiden Seiten einer Medaille sind. Das Kernstück des niederländischen Modells der Qualitätssorge, die die Qualitätsevaluation, -sicherung und gegegenenfalls –verbesserung umfasst, sind die Verfahren der internen und externen Qualitätsevaluation von Lehre (an den Universitäten seit 1988 und an den Fachhochschulen seit 1990) und Forschung (das Verfahren wurde 1993 erstmals in den Fächern Geschichte, Psychologie, Biologie und Maschinenbau erprobt; die erste Runde fachbezogener Forschungsevaluationen erstreckte sich auf den Zeitraum von 1994 bis 1997) unter Hinzuziehung externer Experten.

Vereniging van Universiteiten (1998). Protocol 1998 (Series: Assessment of Research Quality). Utrecht: VSNU.
(ISBN 90-5588-074-4)
http://www.vsnu.nl/servlet/nl.gx.vsnu.client.http.ShowObject?id=97

The Protocol is the basis for the assessments to be carried out in the second VSNU programme for the assessment of the quality of research. It outlines the general principles of the procedure and the requirements for all assessments. The emphasis on context specificity will imply that an assessment in a certain area or descipline will need to be more precisely defined in the specific protocols for these assessments. This context will also include the role of research institutes or research schools. The discipline or area protocols will take into account the rules laid down in this general protocol.
The most important functions of VSNU assessments of the quality of university research will be quality assurance (improvement of university research quality as a result of self-regulation within universities, faculties or research institutes); accountability and the collection of information that can be considered relevant to third parties.

VSNU Reports on the quality assessment of the research, 2nd round (1998-2003):
1993
Werktuigbouwkunde en Maritieme Techniek
Archeologie en Geschiedenis
1994
Filosofie
Biologie
1995
Pedagogical and Educational Sciences
Elektrotechniek
Rechtsgeleerdheid
Ruimtelijke Wetenschappen en Milieukunde
Theologie
Civiele Techniek en Geodesie
Economie
Lucht- en Ruimtevaarttechniek (Aerospace Engineering)
1996
Farmacie
Wiskunde en Informatica
Natuurkunde (Physics)
Politicologie, Bestuurskunde en Communicatiewetenschappen
Scheikunde
Sociologie en Antropologie
Aardwetenschappen
Sterrenkunde
Industrieel ontwerpen
1997
Management Science and Business Administration
Universitaire Lerarenopleiding
1999
Biology
Electrical Engineering, Computer Engineering
Onderzoekschool Arbeid, Welzijn en Sociaal-economisch Bestuur
Psychology
Veterinary and Animal Sciences
2000
Civil Engineering, Geodesy
Environmental Sciences
Geographical Sciences
Maritime Technology

Mechanical Engineering
Pedagogics and Education
Philosophy
Socio-cultural Studies
Theology

Vroeijenstijn, A. I. (1999). External Quality Assessment (EQA) in the Netherlands: The Third Generation 2000-2006 (unpublished paper). Utrecht: VSNU.

In 1988 the Association of Universities in the Netherlands (VSNU) organised the first external programme assessments. Between 1988 and 1993, more than 360 programmes have been assessed and since 1993 many of them have been re-assessed. In 1999 all programmes will be assessed by a second time and in the year 2000 the third cycle will start.
Looking back at ten years EQA in Dutch universities, the conclusion may be that it has been successful and that EQA is rooted deeply in the universities. If we have to characterise the three cycles of quality assessment in the Netherlands, we may say that the 1st cycle (1988-1992) is characterised by the assessment of individual programmes as such. The 2nd cycle (1993-1999) is characterised by the comparative approach and the 3rd cycle (2000-2006) probably will be characterised by more flexibility, accreditation or non-accreditation and international benchmarking.

Vroeijenstijn, A. I. (1995). Improvement and Accountability : Navigating Between Scylla and Charybdis. Guide for External Quality Assessment in Higher Education. London: Jessica Kingsley.

Vught, Frans A. van (1997). The Humboldtian University Under Pressure – New Forms of Quality Review in Western European Higher Education. In: Herbert Altrichter, Michael Schratz & Hans Pechar (Hrsg.), Hochschulen auf dem Prüfstand (48-87). Innsbruck: Studien-Verlag.

Der Beitrag unterscheidet zwei Hauptformen der Qualitätssicherung: Zum einen wird Qualität auf der Ebene von Studienprogrammen überprüft (z. B. Dänemark, Niederlande und – teilweise – Großbritannien); im alternativen Verfahren ist die Grundeinheit der Analyse die einzelne Hochschule (Frankreich sowie – teilweise – Großbritannien).

Westerheijden, Don F. (1999). Where are the quantum jumps in quality assurance? Developments of a decade of research on a heavy particle. Higher Education, 38, 233-254.

The author reviews developments in publications on quality assurance of education in (European) higher education over the last decade. The metaphor of sub-atomic structure is used to order the literature in types of similar publications, moving from those closely related with the practice of quality assurance methods to more theoretical publications. Some seminal publications are highlighted as 'quantum jumps'. The article ends by noting some recent trends in quality assurance at the system and institutional levels, as well as mentioning theoretically interesting developments, notably the emergence of neo-institutional approaches.

Westerheijden, Don F. (1997). A solid base for decisions. Use of the VSNU research evaluations in Dutch universities. Higher Education, 33, 397-413.

To gain insight into the use of the VSNU research quality evaluations (since 1993) in the practice of research and of institutional management in Dutch universities, interviews were held in eight cases evaluated in the first year of this procedure. The main conclusions are that use of these research evaluations is universal, bot 'instrumentally' (in decisions directly based on the judgements) and 'incrementally' (in decision-making processes not directly linked to the evaluation). Underlying this is 'conceptual use': an important change in deans' and rectors' views of their role in managing research, which they now can realise, because the VSNU research evaluations give them, for the first time, solidly legitimate arguments on which to base strategic

decisions. Next to use, other effects can be discerned within universities, pointing to a growing dependency of researchers on managers, necessitating amongst others ever more consciously strategic publication behaviour. Whether quality of research improves in this way, cannot be answered by the study, but certainly it is more difficult for academics not to engage in research.

7. Schweden

Forschungsevaluation

Engwall, Lars (Ed.). (1992). Economics in Sweden: An Evaluation of Swedish Research in Economics. London: Routledge.

The book contains the results of one of the most comprehensive attempts to evaluate research in economics ever undertaken. A team of Swedish and international researchers examined the structure of economics in Sweden and the results it produces. In order to address the various components of the research system, a number of studies were undertaken. Three approaches were used in exploring the work organization. First, a background study was commissioned on the historical traditions of Swedish economic research. Second, the eleven selected institutions were asked to submit material that would ‘provide an adequate picture of the research activities in the department particularly during the 1970s and 1980s'. Third, the international evaluation committee made site visits to all the selected institutions. In order to portray the input and output components of the system three further background studies were commissioned. These concerned the financing of Swedish economic research, the doctoral programmes and the publication practices of Swedish economists. The evaluation report concludes with a summary of good and bad features of the research system in economics in Sweden and recommendations concerning (1) organization of research and (2) postgraduate training.

Guy-Ohlson, Dorothy (1997). International Evaluations of the Swedish Natural Science Research Council (NFR). In OECD (Ed.), The Evaluation of Scientific Research: Selected Experiences (101-106). Paris: OECD/GD(97)194.

Hemlin, Sven (1999). (Dis)Agreement in Peer Review. In Peter Juslin & Henry Montgomery (eds.), Judgment and Decision Making (275-301). Mahwah, NJ: Lawrence Erlbaum.

Peer disagreement is a common phenomenon in science. The chapter focuses on judgments and decisions made by individuals in groups that result in disagreements about a scientific matter. The analysis encompass various kinds of peer review conflicts, whether about a manuscript, a grant proposal, a university department's research, or something else. This research is completed by reviewing relevant literature on peer reviews, group conflicts, group decisions and scientific controversies; and analyzing a peer review conflict on a Ph.D. examination committee.

Hemlin, Sven (1998). Utility evaluation of academic research: six basic propositions. Research Evaluation, 7(3), 159-165.

A review of the literature and three case studies in housing research resulted in a framework of six propositions as to how to assess academic research utility. First, it can be measured directly and indirectly. Secondly, utility is dependent not only on academic research supply of knowledge and technology, but equally importantly on demand from industry and the public sector. Thirdly, research utility should be viewed in short- and long-term perspectives, which makes assessment dependent on time intervals. Fourthly, the framework for evaluating research utility must take into consideration a number of differences with respect to academic research (between applied and basic research, between research areas, between disciplines within an area). Fifthly, a framework must consider the differences in user groups. Finally, the transmission of knowledge from the

academic setting to industry and the public sector is predominantly an interactive process carried out by individuals.

Hemlin, Sven (1996). Research on research evaluation methods. Social Epistemology, 10(2), 209-250.

Hemlin, Sven (1996). Social studies in the humanities: a case study of research conditions and performance in ancient history and classical archaeology, and English. Research Evaluation, 6(1), 53-61.

Research conditions and productivity in two humanistic desciplines are described and compared. A theoretical framework from research evaluation studies which emphasised mediating process factors in conjuction with input factors to explain research output was applied. Results of a questionnaire and interviews showed that the production of publications was generally larger in ancient history and classical archaeology than English. Common features of the two disciplines were a cosmopolitical direction, similar publication profiles and research quality conceptions. Distinguishing factors were related to theoretical, working style and organisational conditions which may have influenced productivity differences.

Hemlin, Sven & Gustafsson, M. (1996). Research production in the Arts and Humanities. A questionnaire study of factors influencing research performance. Scientometrics, 37(3), 417-432.

Jakobsson, S., Oskarsson, C. & Philipson, J. (1996). Indicators of technological activities - comparing educational, patent and R&D statistics in the case of Sweden. Research Policy, 26, 573-585.

Swedish National Board for Industrial and Technical Development (1995). Report of the International Evaluation Committee. NUTEK Info 280-1995. Stockholm.

Swedish Research Council for Engineering Sciences (1997). Peers on Peers: Allocation Policy and Review Procedures at the Swedish Research Council for Engineering Sciences.

Stankiewicz, Rikard (1994). Spin-off companies from universities. Science and Public Policy, 21(2), 99-107.

Strömholm, S. (1999). Peer review - Experience at national and european level. In: European Commission & Austrian Advisory Board for Universities (Eds), Science and the Academic System in Transition: An International Expert Meeting on Evaluation (163-167). Budapest: Akadémiai Kiadó.

Swedish Natural Science Research Council (1999). International Evaluation of Neurobiology. Stockholm: Naturvetenskapliga forskningsrådet (NRF)/Swedish Natural Science Research Council.
http://www.nfr.se/content/publikationer/publikafort.htm

Swedish Natural Science Research Council (1998). Review of the organisation and working methods of the NFR. Stockholm: Naturvetenskapliga forskningsrådet (NRF)/Swedish Natural Science Research Council.
http://www.nfr.se/content/publikationer/publikafort.htm

Swedish Natural Science Research Council (1998). International Evaluation of Exogenic Geochemistry. Stockholm: Naturvetenskapliga forskningsrådet (NRF)/Swedish Natural Science Research Council.
http://www.nfr.se/content/publikationer/publikafort.htm

Swedish Natural Science Research Council (1997). International Evaluation of Mineralogy and Experimental Petrology. Stockholm: Naturvetenskapliga forskningsrådet

(NRF)/Swedish Natural Science Research Council.
http://www.nfr.se/content/publikationer/publikafort.htm

Swedish Natural Science Research Council (1997). International Evaluation of Swedish National Facilities. Stockholm: Naturvetenskapliga forskningsrådet (NRF)/Swedish Natural Science Research Council.
http://www.nfr.se/content/publikationer/publikafort.htm

Swedish Natural Science Research Council (1997). International Evaluation of Fusion Research. Stockholm: Naturvetenskapliga forskningsrådet (NRF)/Swedish Natural Science Research Council.
http://www.nfr.se/content/publikationer/publikafort.htm

Swedish Natural Science Research Council (1997). International Evaluation of Plasma and Space Physics. Stockholm: Naturvetenskapliga forskningsrådet (NRF)/Swedish Natural Science Research Council.
http://www.nfr.se/content/publikationer/publikafort.htm

Swedish Natural Science Research Council (1997). International Evaluation of Analytical Chemistry. Stockholm: Naturvetenskapliga forskningsrådet (NRF)/Swedish Natural Science Research Council.
http://www.nfr.se/content/publikationer/publikafort.htm

Swedish Natural Science Research Council (1995). International Review of the Swedish Research in Mathematical Science. Stockholm: Naturvetenskapliga forskningsrådet (NRF)/Swedish Natural Science Research Council.
http://www.nfr.se/content/publikationer/publikafort.htm

Swedish Natural Science Research Council (1995). International Review of the Swedish Research in the Earth Sciences. Stockholm: Naturvetenskapliga forskningsrådet (NRF)/Swedish Natural Science Research Council.
http://www.nfr.se/content/publikationer/publikafort.htm

Swedish Natural Science Research Council (1995). International Review of the Swedish Research in Biology within the NFR Sphere of Interest. Stockholm: Naturvetenskapliga forskningsrådet (NRF)/Swedish Natural Science Research Council.
http://www.nfr.se/content/publikationer/publikafort.htm

Swedish Natural Science Research Council (1995). International Review of the Swedish Research in Fundamental Chemistry. Stockholm: Naturvetenskapliga forskningsrådet (NRF)/Swedish Natural Science Research Council.
http://www.nfr.se/content/publikationer/publikafort.htm

Wennerås, C. & Wold, A. (1997). Nepotism and sexism in peer-review. Nature, 387, 341-343.

Hochschulevaluation Schweden

Bauer, Marianne & Franke, Sigbrit (1997). Higher Education and Evaluation in Sweden – Changes, Characteristics and Challenges. In: Herbert Altrichter, Michael Schratz & Hans Pechar (Hrsg.), Hochschulen auf dem Prüfstand (134-143). Innsbruck: Studien-Verlag.

In Schweden hat die Hochschulreform von 1993 den Universitäten und regionalen *university colleges* mehr Autonomie verbunden mit der Verpflichtung auf Qualitätssicherung gebracht. Die Autorinnen berichten über die Ergebnisse einer Evaluation, die die Akzeptanz der HochschullehrerInnen für die Grundintentionen der Reform, aber auch die dringende

Notwendigkeit von Fortbildung der Universitätsangehörigen im Hinblick auf Qualitätsentwicklung zeigt.

Engwall, Lars (1997). A Swedish Approach to Quality in Higher Education: The Case of Uppsala University. In John Brennan, Peter de Vries & Ruth Williams (Eds.), Standards and Quality in Higher Education (220-244). London: Jessica Kingsley Publishers.

Högskoleverket (Ed.). (1997). Quality assurance as support for processes of innovation - The Swedish model in cooparative perspective. Stockholm: Högskoleverket (The National Agency for Higher Education), Högskoleverket Studies 1997:1 S.
Högskoleverket Studies in English: http://www.hsv.se/english/studies.html

Sweden, like most other countries, has introduced mechanisms for external quality assurance in its system for higher education during the past ten years. What makes the Swedish approach different from that of most other countries is the emphasis on improvement rather than control. The aim is to support change and renewal; not only to monitor the present. In other words, it has a dynamic perspective rather than a static one.
Four kinds of quality assurance are organised by the National Agency for Higher Education. Two are variants of accreditation, granting rights to award degrees and to establish professorships at non PhD-granting institu-tions. These accreditations have had a profound effect by highlighting the necessary conditions for high quality in higher education institutions. Quality assessments are organised to look at disciplines or particular aspects of the educational system on a national basis, with a focus on suggestions for improvement. Quality audits ask for each institution's conserted efforts to improve its activities. While accreditation is necessary as a minimum standard for public funding, none of the others are directly tied to funding decisions by the government.
The design of quality assurance in Sweden is based on a consensus between the government and the higher education institutions. The latter have accepted both the need to show the outside how public funds are used and the need to revitalise the internal culture. The government, on the other hand, has realised that trust is necessary for a „control system“ to reach beyond the simplest kinds of information on what goes on at such diverse institutions as universities and university colleges, and for the institutions to strive for excellence rather than meeting specified standards.
The improvement-oriented approach is necessary to reach the challenging desired changes in Swedish higher education institutions, which are primarily public agencies, operating under government control; a shift from a slightly inward-looking, rule-obeying culture of bureaucracy to a self-regulating, outward-looking culture of professionalism, where teachers take responsibility or what students learn.

Högskoleverket (1996). The National Quality Audit of Higher Education in Sweden. Stockholm: Högskoleverket (The National Agency for Higher Education), Högskoleverket Reports 1996: 10 R.

Högskoleverket Reports in English: http://www.hsv.se/english/reports.html

In this report the National Swedish Agency for Higher Education presents the basis for its implementation of quality auditing at Swedish universities and colleges, as well as a description of what is evaluated and how the evaluation is to be carried out. A more detailed presentation of the evaluative process and its practical application may be found in two appendices: The National Quality Audit of Higher Education in Sweden: Guidelines for Institutions (Appendix 1), and The National Quality Audit of Higher Education in Sweden: Auditors' Handbook (Appendix 2).

Högskoleverket (1996). Quality Audit of Uppsala University. Stockholm: Högskoleverket (The National Agency for Higher Education), Högskoleverket Reports 1996: 28 R.
Högskoleverket Reports in English: http://www.hsv.se/english/reports.html

Melin, G. (1996). The networking university: A study of a Swedish university using institutional co-authorships as an indicator. Scientometrics, 35(1), 3-13.

Nilsson, Karl-Axel & Näslund, Hans (1996). Towards a Swedish Evaluation and Quality Assurance System in Higher Education. In: Jens-Christian Smeby (ed.), Evaluation of Higher Education in the Nordic Countries (S. 79 – 94). Copenhagen: Nordic Council of Ministers (Nord 1996:6).

Since 1992 representatives from national units responsible for evaluation and Ministers of Education in the Nordic countries (Denmark, Finland, Iceland, Norway and Sweden) have met annually to discuss experiences concerning the evaluation of higher education. During these meetings similarities between the systems and methodologies implemented were noticed. This raised the idea to disseminate Nordic experience to people involved in the evaluation of higher education on a national, institutional and departmental level in the Nordic countries and to an international audience.

The focus of the report „Evaluation of Higher Education in the Nordic Countries" is to examine the systems and methods for evaluation and quality assessment in higher education in the Nordic countries.

According to Nilsson and Näslund (1996) a national system of evaluation of higher education in Sweden is still in a formative stage. Several institutions have adopted a model for local programme or departmental reviews, with self-evaluation and peer review as cardinal elements. These local efforts are basically geared towards internal quality development. On a national level, two or three parallel evaluation systems seem to be evolving. National programme and discipline reviews conducted or initiated by the National Agency will focus on the quality of education, while institutional reviews will aim towards an assessment of the quality system of a whole institution. A third type of evaluation on a national level, which involves a more explicit element of quality control, is the accreditation process that is needed when an institution wants to establish a new type of degree. It is still too early to judge with any certainty how these different elements will work together.

Nybom, T. (1999). Quality assessment and structural change in universities. In: European Commission & Austrian Advisory Board for Universities (Eds), Science and the Academic System in Transition: An International Expert Meeting on Evaluation (43-49). Budapest: Akadémiai Kiadó.

Swedish Research Council for Engineering Sciences (1993). Mechanical Engineering: Doctoral Education and Academic Research in Sweden. Stockholm: Teknikvetenskapliga ForskningsRådet (TFR Evaluation Programme, volume 1).

The doctoral and research programmes in mechanical systems in Swedish universities were reviewed. The review sought the policies and procedures followed by the universities affecting the quality and efficiancy of doctoral education and research in these fields when compared to programmes in leading research universities in the United States.

Swedish Research Council for Engineering Sciences (1995). Applied Mathematics in Sweden for Engineering Sciences. Stockholm: Teknikvetenskapliga ForskningsRådet (TFR Evaluation Programme, volume 2).

Swedish Research Council for Engineering Sciences (1995). Engineering Noise Control: Evaluation of Research and Education in Sweden. Stockholm: Teknikvetenskapliga ForskningsRådet (TFR Evaluation Programme, volume 3).

The Government Bill, The Action Programme against Noise, takes a positive view of increased funding for research on engineering noise control, provided that any increase in funding is preceeded by a scientific evaluation of the research carried out in this area. TFR was therefore commissioned by the Swedish government to carry out an evaluation on the research on engineering noise control in Sweden and to suggest where the main emphasis should be.

Swedish Research Council for Engineering Sciences (1997). Peers on Peers: Allocation Policy and Review Procedures at the Swedish Research Council for Engineering Sciences. Stockholm: Teknikvetenskapliga ForskningsRådet (TFR Evaluation Programme, volume 4).

The report consists of two parts: Part 1 reflects the impressions of an international review panel after reading background information about TFR and after conducting their own interviews with selected researchers and meeting a number of representatives from Swedish universities, other funding agencies, and industry. Part 2 is the result of the evaluative investigation on funding policy, researchers' attitudes and Council procedures.

8. Finnland

Forschungsevaluation

Commissioner: Academy of Finland

Academy of Finland (1998). The State and Quality of Scientific Research in Finland. Helsinki: Edita (Publications of the Academy of Finland, 2/98, Evaluation Report).

Academy of Finland (1998). Report of the working group on indicators. Helsinki: Edita (Publications of the Academy of Finland, 9/98).

In the beginning of 1997 the Academy of Finland set up a working group to develop the indicators illustrating the Academy's performance and the basis of the performance data provided in the annual review. The working group studied the principles and development needs of the process applied in planning the Academy's operations and the role of management by results in that process. In addition, it made proposals concerning targets derived from the Academy's operating principles as well as indicators used to evaluate the attainment of these targets.

Academy of Finland (1997). Evaluation of Electronics Research in Finland. Helsinki: Edita.

Academy of Finland (1997). Molecular Biology and Biotechnology Research in Finland. EMBO Evaluation Report. Turku: Koteva Oy.

Financing and development programmes for biotechnology and molecular biology research have been implemented since 1988. The success of the programmes has been evaluated by the joint action of The Academy of Finland and The European Molecular Biology Organisation (EMBO).

Academy of Finland (1996). Psychiatric Research in Finland in 1995. A Peer Review Report for the Academy of Finland. Helsinki.

Academy of Finland (1996). Evaluation of the Finnish Research Programme on Climate Change. Helsinki.

Academy of Finland (1995). Evaluation of the National Public Health Institute of Finland – Report of the Evaluation Panel. Helsinki: Edita (Publications of the Academy of Finland, 9/95).

In 1994, the Academy of Finland, in response to a proposal put to it by the national Public Health Institute of Finland (KTL), decided to carry out an evaluation of the research activities of the KTL. The terms of reference were: (1.) to evaluate the public health function, strategic importance, scientific merit, and effectiveness in the use of resources; (2.) to evaluate the strategic importance and scientific merit of the proposed future programme for the KTL in relation to its objectives; and (3.) to provide the Academy of Finland and the Ministry of Social Affairs and Health with an evaluation of the current and planned work and make recommendation on the strategic development, organisation and resourcing of the KTL.

Commissioner: Bank of Finland

Bank of Finland (1999). An Evaluation of the Research Activities of the Bank of Finland. Helsinki.

Commissioner: Ministry of Agriculture and Forestry

Ministry of Agriculture and Forestry (1999). International Evaluation of the Finnish Game and Fisheries Research Institute – Report of the Evaluation Group. Helsinki.

Ministry of Agriculture and Forestry (1998). Evaluation of the Finnish forest research institute Metla – Report of the Evaluation Panel. Helsinki.

Ministry of Agriculture and Forestry (1998). Evaluation of the National Veterinary and Food Research Institute EELA – Report of the Evaluation Group. Helsinki.

Ministry of Agriculture and Forestry (1998). Getting Ready for the Next Century. Evaluation of the Finnish Geodetic Institute. Helsinki.

Ministry of Agriculture and Forestry (1996). Evaluation of the Agricultural Research Centre of Finland (MTT) – Report of the Evaluation Panel. Helsinki.

Commissioner: Ministry of Education

Ministry of Education (1999). From the Cherry Orchard to the Future. Evaluation of the Theatre Academy of Finland. Helsinki.

Ministry of Education (1998). Three Finnish Universities in the International Perspective. CRE Institutional Review of Helsinki University of Technology, Tampere University of Technology and Åbo Akademi University. Helsinki.

Ministry of Education (1995). Evaluation of the Sibelius Academy – Report of the External Evaluation Group. Helsinki.

Blume, Stuart, Heløe, Leif Arne, Larsen, Peder Olesen & Posner, Michael V. (1993). The Academy of Finland – An International Evaluation 1992. Helsinki: Ministry of Education.

Commissioner: Ministry of Justice

Ministry of Justice (1997). Oikeuspoliittinen tutkimuslaitos – arviointiraportti. Evaluation of the National Research Institute of Legal Policy. Helsinki.

Commissioner: Ministry of Social Affairs and Health

Ministry of Social Affairs and Health (1999). International Evaluation of the National Research and Development Center for Welfare and Health. Helsinki.

Ministry of Social Affairs and Health (1995). Työterveyslaitos – Investment in Health. The Scientific and Functional Evaluation of the Finnish Institute of Occupational Health. Helsinki.

Commissioner: Ministry of the Environment

Ministry of the Environment (1998). Futures for FEI. International Evaluation of the Finnish Environment Institute. Helsinki.

Commissioner: Ministry of Trade and Industry

Ministry of Trade and Industry (1998). Evaluation of the Technological and Industrial Benefits of Finnish Space Programmes. Helsinki.

Ministry of Trade and Industry (1998). Innovation and Invention in Finland Strategies for Networking – An International Evaluation. Helsinki.

Guillaume, Henri & Zegveld, Walter (1995). The Technology Development Centre of Finland (TEKES) – An International Evaluation. Helsinki: Finnish Ministry of Trade and Industry, Publications 5/1995.

Tomner, Sigvard & Zegveld, Walter (1993). The Technical Research Centre of Finland (VTT) – An International Evaluation. Helsinki: Finnish Ministry of Trade and Industry, Publications 2/1993.

Commissioner: The Technology Development Centre of Finland (TEKES)

Technology Development Centre of Finland (2000). Nanotechnology Research Programme, 1997-1999. Evaluation Report. Helsinki: TEKES (Technology Programme Report, 11/2000).

Technology Development Centre of Finland (1998). Technology Strategy Consulting for SMEs. Evaluation Report. Helsinki: TEKES (Technology Programme Report 13/98).

Technology Development Centre of Finland (1998). Digital Media in Finland. Evaluation Report. Helsinki: TEKES (Technology Programme Report 12/98).

Berg, Pekka & Lindberg, Ralf (1997). Assessment and Decision Making for R&D Programmes. Helsinki: TEKES (Technology Programme Report 16/97).

Buchwald, Stephe L., Danheiser, Rick L. & Schrock, Richard R. (1996). Synthesis Technology Programme, 1992-1996. Evaluation Report. Helsinki: TEKES (Technology Programme Report 11/96).

Klipstein, David H. & McRae, Gregory J. (1996). Evaluation Report of the Process Technology Programme. Helsinki: TEKES (Technology Programme Report 5/96).

Langer, Robert S., Cooney, Charles L. & Brain, Joseph D. (1996). Pharmaceutical Technology Programme, 1989-1994. Evaluation Report. Helsinki: TEKES (Technology Programme Report 12/96).

Commissioner: The Technical Research Centre of Finland (VTT)

Technical Research Centre of Finland (1998). VTT Chemical Technology Evaluation Report. Helsinki.

Technical Research Centre of Finland (1998). VTT Electronics Evaluation Report. Helsinki.

Technical Research Centre of Finland (1997). VTT Information Technology Evaluation Report. Helsinki.

Technical Research Centre of Finland (1996). VTT Biotechnology and Food Research Evaluation Report. Helsinki.

Technical Research Centre of Finland (1996). VTT Automation Evaluation Report. Helsinki.

Technical Research Centre of Finland (1995). Research Programme on Molecular Modelling. Evaluation Report. Helsinki.

Helander, Elisabeth (1995). Evaluation Activities in the Nordic Countries. Scientometrics, 34(3), 391-400.

Evaluations of whole areas of research started in the Nordic countries in the early 1980's. There has been extensive experiences with evaluations in the Nordic Countries. The paper gives a brief overview of work related to: evaluation of research fields, bibliometric studies, evaluations of

research programmes, performance of research institutes, evaluation of bodies supporting research, evaluation of universities, indicators and databases.

Kaukonen, Erkki (1997). Evaluation of Scientific Research in Finland. In OECD (Ed.), The Evaluation of Scientific Research: Selected Experiences (12-26). Paris: OECD/GD(97)194.

Lindholm-Romantschuk, Ylva & Warner, Julian (1996). The role of monographs in scholarly communication: an empirical study of Philosophy, Sociology and Economics. Journal of Documentation, 52(4), 389-404.

Luukkonen, Terttu (1999). Finnish (Nordic) Culture. In: Susanne Bührer & Stefan Kuhlmann (Eds.), Evaluation of Science and Technology in the new Europe. Proceedings of an International Conference on 7 and 8 June 1999, Berlin (49-54). Bonn / Bruxelles: Federal Ministry of Education and Research / European Commission.

Luukkonen, Terttu (1997). Quantitative Techniques in Evaluation in Western Europe. In Mark S. Frankel & Jane Cave (Eds.), Evaluating Science and Scientists: An East-West Dialogue on Research Evaluation in Post-Communist Europe (117-131). Budapest: Central European University Press.

The author provides an overview of the growing use, in several West European countries, of quantitative techniques to complement other types of information when evaluating scientific research. She examines the role of quantitative indicators in the evaluation of individual scientists, research groups, research institutions, scientific fields, research programs and national research performance, and cautions that interpretation of citations 'as an indication of quality' is problematic.

Luukkonen, Terttu (1997). The Increasing Professionalisation of the Evaluation of Mission-oriented Research in Finland: Implications for the Evaluation Process. In OECD (Ed.), Policy Evaluation in Innovation and Technology Towards Best Practices (347-356). Paris: OECD.
http://www.oecd.org/dsti/sti/stat-ana/prod/luukkonen.pdf

Luukkonen, Terttu (1995). The impacts of research field evaluations on research practice. Research Policy, 24, 349-365.

Luukkonen, Terttu (1991). Citation indicators and peer review: their time-scales, criteria of evaluation, and biases. Research Evaluation, 1(1), 21-30.

Citations have been increasingly used in research evaluation in recent years. The paper assesses citations as a measure of performance by comparing them with peer judgment. It considers the differences of these two methods, and pays attention to some factors other than quality which potentially affect the accumulation of citations and the relative comparisons of research groups and university departments - orientation in basic or applied research and the rate of self-citations. The comparisons between citation counts and peer judgment produced inconsistent results. Self-citations did not at all affect the relative comparisons based on citations, and research orientation had less influence than expected.

Luukkonen, Terttu & Ståhle, Bertel (1990). Quality Evaluations in the Management of Basic and Applied Research. Research Policy, 19(4), 357-368.

Numminen, Sirkka & Hämäläinen, Olli (1995). Evaluation of TEKES funding for industrial R&D. An empirical study of 601 industrial R&D projects funded by the Technology Development Centre of Finland (TEKES). Espoo 1995. Technical Research Centre of Finland, VTT Tiedotteita - Meddelanden - Research Notes 1661.

Hochschulevaluation Finnland

Davies, John, Lindström, Caj-Gunnar & Schutte, Frits (1999). Five Years of Development: Follow-Up Evaluation of the University of Oulu. Helsinki: Oy Edita Ab (Publications of the Higher Education Evaluation Council 7:1999).
http://www.minedu.fi/asiant/kka.html

From when it was launched in the early 1990s, the practice of Finnish institutional evaluations has served one purpose: to develop the universities subject to it. The first two evaluations, conducted at the universities of Oulu and Jyväskylä, piloted a programme of the Ministry of Education, aimed at establishing an evaluation procedure that could be applied in all Finnish institutions of higher education. There were three phases in the first evaluations: self-evaluation, an external evaluation, and a published report. This model, already widely in use abroad, became the standard also for Finnish institutional evaluations. There is, however, a distinctive feature in the Finnish evaluation practice: the evaluation are tailored to the needs of the universities. Now that the first round of institutional evaluations is almost completed, the Higher Education Evaluation Council has offered the institutions an opportunity to a follow-up evaluation. In addition to the repercussions of internal evaluation, several external factors have had an impact on the development of the universities: the national evaluations of three fields of education (sciences, humanities and education), coniderable cutbacks in university budgets, extensive deregulation, and the introduction of performance-based steering system.

Finnish Higher Education Evaluation Council (1999). Evaluation of higher education – the first four years. Helsinki. (ISBN 951-98446-0-0)

Many European countries have a national, state-established organisation to conduct evaluations of higher education. The oldest have been established for approximately ten years and the most recent ones, such as that in Finland, for only a couple of years. Although the size, policies and administration of such organisations vary, even within the Nordic countries, all have the same objective: to improve higher education. Established in 1996 and funded by the Ministry of Education, the Finnish Higher Education Evaluation Council is an expert body which organises and implements evaluations. It functions independently of the educational administration and institutions of higher education. All Finnish universities will have been evaluated by the end of 2000. Since 1999, the Evaluation Council has been in charge of the evaluation and accreditation of professional courses.

Goddard, J., Moses, I., Teichler, U., Virtanen, I. & West, P. (2000). External Engangement and Institutional Adjustment: An Evaluation of the University of Turku. Helsinki: Oy Edita Ab (Publications of the Higher Education Evaluation Council 3:2000).
http://www.minedu.fi/asiant/kka.html

What generates a university's external impact? Is something other than research and education needed? In the report the international Peer Review Team of the University of Turku considers both the external impacts of the University and the mechanisms needed for interaction with external stakeholders.

Further Publications of The Finnish Higher Education Evaluation Council in English:
7:1998. Pilot Audit of Quality Work in Kajaani, Turku, Lahti and Häme Polytechnics.
8:1998. Towards the Responsive University. The Regional Role of Eastern Finland Universities.
9:1998. Programme Evaluation of Industrial Management and Engineering.
10:1998. Quality Label? EQUIS Evaluation Report. Helsinki School of Economics and Business Administration.

11:1998. Three Finnish Universities in the International Perspective. CRE Institutional Review of Helsinki University of Technology, Tampere University of Technology and Åbo Akademi University.
3:1999. Strategies for the Future. Evaluation of University of Art and Design Helsinki UIAH.

Hämäläinen, Kauko & Moitus, S. (1999). High-quality Education as the Criterion for University Funding in Finland. Quality in Higher Education, 5(1).

Lindqvist, O. V. (1999). Quality assessment and structural change in universities. In: European Commission & Austrian Advisory Board for Universities (Eds.), Science and the Academic System in Transition: An International Expert Meeting on Evaluation (39-42). Budapest: Akadémiai Kiadó.

Puukka, Jaana (2000). External impact of the University of Turku. Self-Evaluation Report. University of Turku, Rector's Office (Publications 1/2000).

Increasing university self-regulation, and changes in the funding base and in the higher education environment are all factors that call for greater external involvement from the universities. In recent years, the University of Turku has become more active in serving the needs of the region and society in general, and has opened up more than before to external stakeholders. In order to improve its institutional capacity to respond to the external needs, the University decided to make external impact the subject of its self-evaluation.

9. Frankreich

Forschungsevaluation

Barré, Rémi (1999). French Culture. In: Susanne Bührer & Stefan Kuhlmann (Eds.), Evaluation of Science and Technology in the new Europe. Proceedings of an International Conference on 7 and 8 June 1999, Berlin (45-47). Bonn / Bruxelles: Federal Ministry of Education and Research / European Commission.

Barré, Rémi, Laville, F., Teixeira, N. & Zitt. M. (1995). L'Observatoire des sciences et des techniques: activités - définition - méthodologie. http://www.info.unicaen.fr/bnum/jelec/Solaris/d02/2barre.html

Bauin, S., Michelet, B., Schweighofer, M. G. & Vermeulin, P. (1991). Using bibliometrics in strategic analysis: „Understanding chemical reactions" at the CNRS. Scientometrics, 22(1), 113-137.

Callon, Michel (1998). Strategic Management of Research and Technology. Washington, D. C.: Brookings Institute. (ISBN: 1902282027)

Callon, Michel, Larédo, Philippe & Mustard, Philippe (1995). La gestion stratégique de la recherche et de la technologie. Paris: Economica (ISBN 2-7178-2853-2).

The book contains contributions of 28 experts from England, France, Germany, the Netherlands, Finland and Switzerland. It offers an excellent survey of the state of the art in research evaluation, as reflected in the accumulated experience within the European Union over the last 15 years.

Callon, Michel, Larédo, Philippe & Rabeharisoa, V. (1992).The Management and Evaluation of Technological Programs and the Dynamics of Techno-Economic Networks: The Case of the Agence Francaise de la Maitrise de I'Energie (AFME). Research Policy, 21, 215-236.

Callon, Michel, Maurice, Marc & Musselin, Christine (Eds.). (1996). Recherche Scientifique, Innovation Techniques et Politiques Publiques. Sociologie du Travail, 38(3), 253-425.

Centre National de la Recherche Scientifique (1995). Audit du Comité national de la recherche scientifique. Paris: CNRS éditions.

Comité national d'évaluation (1993). Universités: la recherche des équilibres. Paris: La Documentation Francaise.

Comité national d'évaluation (1989). Priorités pour l'Université. Rapport de fin de mandat. Paris: La Documentation Francaise.

Comité national d'évaluation de la recherche (1993a). L'Evaluation de la Recherche: Réflexions et Partiques. Paris: La Documentation Francaise.

Comité national d'évaluation de la recherche (1993b). Un autre regard sur la recherche – Sept evaluations 1990-1993. Paris: La Documentation Francaise.

Comité national d'Eéaluation de la recherche (1994). Réflexions sur l'appareil national de recherche et de dévelopement technologique. Rapport d'activité au Président de la République.

Comité national d'évaluation de la recherche (1996). L'évaluation de la recherche: un enjeu capital. Rapport d'activité au Président de la République.

Larédo, Philippe (1997). Evaluation in France: A Decade of Experience. In OECD (Ed.), Policy Evaluation in Innovation and Technology Towards Best Practices (Chapter 24). Paris: OECD.
http://www.oecd.org/dsti/sti/stat-ana/prod/evaluation.htm

Larédo, Philippe (1995). The Impact of Community Research Programmes in France. Paris: Les Presses de l'Ecole des Mines.

The French study emphasises the ever increasing significance of the emergence of multi-national research networks on the basis of the European cooperative research projects. An important element is the development that Larédo calls „hybridisation", that is, the collaboration of partners from different institutional backgrounds (public research laboratories, big companies, small companies, universities).

Larédo, Philippe & Mustar, Philippe (1995). France, the guarantor model and the institutionalisation of evaluation. Research Evaluation, 5(1), 11-21.

The French evaluation scence highlights a particular configuration - the guarantor model. Historically, evaluation has played a major role in the recruitment and careers of researchers, and, since the 1960s, advisory committees have been responsible for evaluating the French R&D situation and policy. At the beginning of the 1980s two independent bodies - CNE (Comité National d'Evaluation) and CNER (Comité National d'Evaluation de la Recherche) - were given the responsibility of systematically and periodically evaluating all French research operators - research institutions, national programmes and agencies, universities, and 'procedures' (such as the research tax credit). The review of their first years of experience highlights three lessons: they were able to produce credible evaluations (a key issue for such exercises), they have had significant effects, both direct and indirect, on research operators, but the question of the uptake of their results at the policy level remains open.

Lautman, Jacques (1992). Evaluation at the CNRS. In: Andrea Orsi Battaglini, Michel Lesage & Francesco Merloni (Eds.), Scientific Research in France – Problems in Administration, Evaluation and Planning (79-107). Baden-Baden: Nomos Verlagsgesellschaft.

The main evaluating body within the CNRS (France's National Council for Scientific Research) is the National Committee of Scientific Research (Comité National de la Recherche Scientifique) which amounts to a parliament of basic research in France, being charged with three objectives: (1.) evaluating individuals, candidates for research positions and reseachers in the midst of their careers for the purpose of eventual promotion; (2.) evaluating laboratory programs associated with the CNRS, or programs that wish to be associated; and (3.) drawing up reports at four-years intervals on the current scientific situation and the prospects for the future.

Ledoux, M. J. (1999). Measuring the impact of the EU Framework Programme. In: European Commission & Austrian Advisory Board for Universities (Eds), Science and the Academic System in Transition: An International Expert Meeting on Evaluation (177-178). Budapest: Akadémiai Kiadó.

In the past twenty years the BETA has performed many evaluation excercises on different R&D programmes. The adopted quantitative microeconomic approach is concerned with direct and indirect economic effects generated according to the nature of the programme participants (big firms, SMEs, research and academic organizations), and with the impact of S&T programmes on employment. The method of evaluation provides indications on the implementation of S&T programmes on the micro level, and - at the macro level - on European competitivity and European conhesion.

Le Minor, Sylvaine & Dostatni, Paulette (1991). A bibliometric study of the publications of the French National Institute for Health and Medical Research (INSERM). Scientometrics, 22(1), 41-63.

The study helps to situate the organization's written production in a national and an international context and, in particular, to trace the 'profile' of the Institute's researchers and the impact of the journals used in their publications.

Noyer, Jean-Max (1995). Scientométrie, infométrie : pourquoi nous intéressent-elles? http://www.info.unicaen.fr/bnum/jelec/Solaris/d02/2noyer_1.html

Observatoire des Sciences et des Techniques (1999). Science & Technologie Indicateurs 2000. Paris: Éditions Économica.

Chapitre 6: L'Union européenne, les Etats-Unis et le Japon
Le financement et l'exécution de la RD de l'Union européenne, des Etats-Unis et du Japon
La production scientifique: la mesure par les publications
La technologie et la compétitivité industrielle
Les coopérations, réseaux et échanges internationaux

Quoniam, L., Rostaing, H., Boutin, E. & Dou, H. (1995). Treating bibliometric indicators with caution: their dependence on the source database. Research Evaluation, 5(3), 177-181.

Nowadays, with computer-supported analysis of databases, constructing bibliometric or scientometric indicators may be considered easy. The problem is more to verify the accuracy of the global analysis, including the sampling of data. The global coherence of an analysis depends on the adequacy of all the steps. Using on-line databases, an experiment was designed to demonstrate this. Keeping the same protocol for data collection, the same indicators are used over the various samples. The results from three separate databases are profoundly different.

Papon, P. (1999). The role of national agencies in evaluation. European Commission & Austrian Advisory Board for Universities (Eds.), Science and the Academic System in Transition: An International Expert Meeting on Evaluation (63-71). Budapest: Akadémiai Kiadó.

Papon, Pierre (1996). A New Context for Scientific Expertise? Some Lessons from the French Experience. Minerva, 34, 151-160.

Scientific research in developed countries has several goals: the production of new knowledge, innovation within the framework of corporate policies, contributions to strategic public programmes in, for example, defence, space and nuclear energy, and production of the knowledge needed for public policy-making in social areas such as health, environment and transport. In the majority of industrialised countries, 50 to 80 per cent of scientific and technological activity now has economic, social and strategic goals.

Papon, Pierre (1988). Science and Technology Policy in France: 1981-1986. Minerva, 26(4), 493-511.

The reforms along the lines of the loi d'orientation et de programmation, which was enacted in 1982, lowered the barriers separating the various scientific institutions from each other and those scientific institutions and industrial enterprises. Joint ventures and programmes were undertaken by the various partners working together. Within institutions interdisciplinary programmes or projects were started in materials science, new sources of energy, etc. The mobilisation programmes furthered the co-operation between disciplines and between science and technology and industry. The development of the relationship between governmental laboratories and industry was certainly one of the most important achievements of the new science and technology policy.

Schweighofer, Marie-Gabrielle (1997). Research Evaluation at the Centre National de la Recherche Scientifique (CNRS) in France. In OECD (Ed.), The Evaluation of Scientific Research: Selected Experiences (74-82). Paris: OECD/GD(97)194.

Sevin, Jacques & Bauin, Serge (1999). Bibliometric Evaluation of the scientific production of the CNRS (French National Centre for Scientific Research): Life and Physical Sciences, 1986-1996.
http://www.cnrs.fr/DSP/doc/bib99.pdf%20

The bibliometric evaluation of the CNRS' scientific production is clearly worthwhile. Such an evaluation is in any case necessary, regardless of the fact that the Parliament regularly requests figures from organizations, in order to prepare the law on finance. Indeed, an organization devoted to basic research must be capable of providing a demonstration of its results. Such is the aim of the bibliometry specialists of several organizations, among which the CNRS, via the UNIPS (Unité d'indicateurs de politique scientifique). The bibliometric studies concern the life sciences and physical sciences, and the results presented here are only a sample of the indicators that can be developed on the basis of available data.

Zitt, M. & Teixeira, N. (1996). Science macro-indicators: Some aspects of OST experience. Scientometrics, 35(2), 209-222.

Hochschulevaluation Frankreich

Chevaillier, Thierry (1995). Quality assessment in the French higher education system. Dijon: Institute for the Management of Education, Université de Bourgogne.

The major change in higher education has been the introduction of contractual agreements negotiated and signed by each university and the ministry of education. It is part of a larger reform introduced across the whole of public administration in the spirit of the „Management by objective" doctrine. In the field of higher education, the first contracts were signed in 1991.
A reason frequently quoted for explaining the limited extent of quality assessment approaches in France is the lack of reliable data. Universities use accounting and information systems which are ill suited as tools for observation and decision. They were conceived two decades ago as instruments of control by a highly centralised ministry. The main obstacle to quality evaluation in France is the long tradition of centralisation and legalistic attitude to education (and higher education in particular), by which the government is considered responsible for the provision of a uniform and free education to all.

Comité National d'Évaluation (1995). Evolution des Universités, Dynamique de l'Évaluation: Rapport au Président de la République 1985-95. Paris: La Documentation Francaise.

Malicet, Danielle Potocki (1997). Evaluation and Self-evaluation in French Universities. European Journal of Education, 32(2), 165-174.

The evaluation of universities is part of the evaluation of public institutions and public policies, i. a. it is part of the State and Public Service modernisation policy launched by the Rocard government and the circular of 23 February 1989, imposing evaluation as the duty of any administration. Although there is no real hostility, the lack of political will which is clearly asserted by both the central authority and the universities, the lack of external sanction mechanisms for the results, and a lack of coordination in the allocation of resources, as wll as financial and orgaisational weaknesses on the part of the existing services make it very difficult to implement evaluation and self-evaluation in French universities.

Neave, Guy (1996). The evaluation of the higher education system in France. In Robert Cowen (Ed.). World Yearbook of Education 1996: The Evaluation of Higher Education Systems (66-81). London: Kogan Page.

Ottenwaelter, Marie-Odile (1997). Evaluation à la Francaise: The CNE. In John Brennan, Peter de Vries & Ruth Williams (Eds.), Standards and Quality in Higher Education (78-86). London: Jessica Kingsley Publishers.

The Comité National d'Évaluation des établissements publics à caractère scientifique, culturel et professionnel (CNE) was created in 1984. The CNE enjoys full administrative autonomy: it reports directly to the President of the Republic. The CNE is involved in three types of activities. First, it carries out institutional evaluations, that is, evaluations of universities and écoles. In 1996, it had completed the evaluation of all French universities and that of about 20 schools (it has published over 100 reports. It is possible to consult the reports on the Internet: http://www-cne.mesr.fr/). The CNE has started the evaluation of universities that had already been evaluated (second-round evaluations) to assess the implementation of its initial recommendations and to measure change; it plans to conduct such evaluations on a five-year cycle. Second, the CNE undertakes cross-cutting and comparative evaluations which assess either a specific discipline (e.g., geography) or a degree (e.g., magistères). And third, based on knowledge acquired through the institutional and disciplinary evaluations, the CNE studies the missions of universities.

Staropoli, André (1992). Evaluating Universities. In: Andrea Orsi Battaglini, Michel Lesage & Francesco Merloni (Eds.), Scientific Research in France – Problems in Administration, Evaluation and Planning (57-78). Baden-Baden: Nomos Verlagsgesellschaft.

Within the framework of institutional evaluation in France, judgement by peers is the rule, with this process being backed by a solid foundation of quantitative data (e. g. research funds, publications, students working on doctoral theses, quality and results, i. a. prizes and awards, patents, aiding in the establishment of business enterprises.

10. Norwegen

Forschungsevaluation

Brofoss, Karl Erik (1998). The Research Council of Norway's use of research evaluation: an assessment of research evaluation as a strategic tool. Research Evaluation, 7(3), 134-140.

The Research Council of Norway (RCN), established at the beginning of 1993, was given responsibility for implementing the Government's research policy and being its prime research policy adviser. The paper examines the policy context within which RCN operates, the evaluation portfolio and the utilisation profile, and analyses the interaction among different types of use of evaluation results. The study concentrated on whether or not the evaluation portfolio has a strategic orientation. It concludes that the Council has not used the full potential of research evaluation as a strategic tool. To do so, it must allow a more systematic evaluation approach to permeate the organisation whereby a common frame of reference regarding choice of evaluation subjects, choice of strategic focus, and use of fundamental concepts is established.

Brofoss, Karl Erik (1993). Government approaches to evaluations. Research Evaluation, 3(3), 187-195.

An overview is given of research evaluation in Norway with special emphasis on research institutes. These evaluation efforts are best characterised as very heterogeneous. There is so far no consensus on which aspects of an institute's activities should be included in an evaluation. Some evaluations concentrate on organisational aspects, others solely on research quality, and others on the interaction between research establishments and policy institutions. There is a need for a more integrated approach in which each of these aspects is part of the evaluation and they can mutually benefit from each other.

Research Council of Norway (1995). Evaluation procedures. Dosky.dok.no 3-7-3.

Skoie, H. (1999). Bibliometrics - Some warnings from the North. In: European Commission & Austrian Advisory Board for Universities (Eds), Science and the Academic System in Transition: An International Expert Meeting on Evaluation (105-109). Budapest: Akadémiai Kiadó.

11. Japan

Forschungsevaluation

Barker, B. G. (1996). Japan: A Science Profile. London: The British Council & Department of Trade and Industry. ISBN: 0-863-55326-7.

Gonda, Kinji & Kakizaki, Fumihiko (1995). Research, technology and development evaluation: Developments in Japan. Scientometrics, 34(3), 375-389.

In this paper, two cases of in-house RTD evaluation are described: (a) in Riken, which is a semi-public research corporation of the Science and Technology Agency, and (b) in regional public research institutes. Furthermore, RTD evaluation from the view point of policy assessment of governmental science and technology policy is discussed through analysis of data obtained by the survey of research activities in regional public research institutes. It can be concluded that developments and introduction of RTD evaluation as a new management system in these instituts is improving the research environment and advancing the quality of research.

Irvine, John (1988). Evaluating Applied Research: Lessons from Japan. London: Printer Publishers.

The report presents the findings of a study on the evaluation of applied research in Japan commissioned by the UK Department of Trade and Industry. Despite the central importance of Japan in the world R&D arena, there is still very little English-language material available on how its research system operates. Over twenty Japanese organizations responsible for commissioning or conducting evaluation of research were visited. The questions addressed focused on the methods and techniques employed for evaluating applied research; the extent to which they are used for routine monitoring, mid-term and ex-post evaluation; the strengths and weaknesses of different approaches to evaluation; the objectives of assessment activities; details relating to the planning, organization and timing of evaluation; and allocation of responsibility for overseeing and executing assessment activities. The author concludes, that Japan has developed a highly effective system for planning, managing and evaluating research intended to make incremental contributions to science and technology. Japan's major aim now is to develop a more appropriate environment for achieving similar levels of success in creative longer-term research.

Kameoka, A. (1995). Evaluating research projects at Toshiba. Designing a conceptual framework of evaluating research and technology development (RTD) programs. Scientometrics, 34(3), 427-439.

Mitsuma, Hidehiko & Hirota, Toshiro (1993). Analyses of National Science and Technology Programs (Interim Report) - Framework of the Research and Preliminary Analyses of Activities of National R&D Organizations and Cooperation of Government-Industry-Academia (Chapter 3.2.5.2: Pre-evaluation of Research Themes). Tokyo: National Institute of Science and Technology Policy (NISTEP), 1st Theory-Oriented Research Group 1993.1., Report No. 26.
http://www.nistep.go.jp/achiev/fulltx-e/report26e/report26e.html

National Institute of Science and Technology Policy (1999). Quantitaive Analysis on Life Science Research in Japan. Tokyo: National Institute of Science and Technology Policy (NISTEP), NISTEP Policy Study No. 4.

National Institute of Science and Technology Policy (1999). Assessment for the Effects of R&D Policy on Economic Growth (Interim Report). Tokyo: National Institute of Science and Technology Policy (NISTEP), NISTEP Report No. 64.

National Institute of Science and Technology Policy (1999). Research Planning Process in National Research Institutes. Tokyo: National Institute of Science and Technology Policy (NISTEP), NISTEP Research Material & Data Report No. 63.

National Institute of Science and Technology Policy (1998). Research Assessment in the United Kingdom - Public Research funding based on „Value for Money“ and „Selectivity“. Tokyo: National Institute of Science and Technology Policy (NISTEP), NISTEP Research Material & Data Report No. 54.

National Institute of Science and Technology Policy (1997). Science and Technology Indicators: 1997 - A Systematic Analysis of Science and Technology Activities in Japan. Tokyo: National Institute of Science and Technology Policy (NISTEP), NISTEP Report No. 50.

National Institute of Science and Technology Policy (1996). How Much R&D Is Needed to Achieve a Targeted Rate of Economic Growth?. Tokyo: National Institute of Science and Technology Policy (NISTEP), NISTEP Research Material & Data Report No. 44.

National Institute of Science and Technology Policy (1994). Characteristics of Excellent Researchers and Their Research Activities in Japan. Tokyo: National Institute of Science and Technology Policy (NISTEP), NISTEP Research Material & Data Report No. 38.

NISTEP Publications:
http://www.nistep.go.jp/achiev/achiev.html

Science and Technology Agency (1999). National Report on R&D Evaluation. Tokyo: Science and Technology Agency; Science and Technology Policy Bureau; Planning and Evaluation Division; Office of Evaluation.
http://www.sta.go.jp/policy/seisaku/e9902_3.html

With respect to the evaluation of research and development (R&D) practiced on the basis of the „National Guideline on the Method of Evaluation for Government R&D“ (decided by the Prime Minister in July, 1998), the Science and Technology Agency (STA) summarized the efforts made by ministries and agencies concerned and published a report of the efforts in a lump as the „National Report on R&D Evaluation“.
The report mainly introduces the making-up of evaluation systems and practical evaluation conditions of ministries and agencies concerned during the period from the decision of the above-mentioned guidelines to August last year, in 2 parts (Part I: General Discussions; and Part 2: Efforts of Evaluation by Ministries and Agencies Concerned).
The ministries and agencies have made manuals (http://www.sta.go.jp/policy/seisaku/e9802_4.html) and the like for the practice of the evaluation and have been proceeding to regularization of the evaluation, but since the efforts are only at their very beginning, it is necessary to solve problems in the practice of the evaluation and try to improve the evaluation to fix its course, the report suggests.

Science and Technology Agency (1998). National Guideline on the Method of Evaluation for Government R&D. Tokyo: Science and Technology Agency; Science and Technology Policy Bureau; Planning and Evaluation Division; Office of Evaluation.
http://www.sta.go.jp/shimon/cst/hyoka/ENGLISH.HTM

CONTENTS

Science and Technology Agency (1998). Inter-Ministerial Group Formed to Promote the Evaluation of Government Research and Development. Tokyo: Science and Technology Agency; Science and Technology Policy Bureau; Planning and Evaluation Division; Office of Evaluation.
http://www.sta.go.jp/policy/seisaku/e9803_9.html

To promote effectively the government research and development (R&D), a „inter-ministerial group for evaluation of government R&D“ composed of 15 ministries and agencies was formed on January 29.
The inter-ministerial group was organized to promote effective R&D evaluation within the government as a whole in compliance with the „National Guideline on the Method of Evaluation for Government R&D“ (Prime Minister's Decision in August, 1997). The group will examine the collaboration and cooperation involved in evaluating R&D and will prepare a report called „Annual Report of the evaluation of Government R&D (tentative name)“ to present the evaluations

conducted by ministries, agencies and research institutes in a single document understandable by the public.
The group is composed of division-directors or similar officials from the National Police Agency, Hokkaido Development Agency, Defense Agency, Science and Technology Agency (STA), Environment Agency, Ministry of Finance, Ministry of Education, Science, Sports and Culture (Monbusho), Ministry of Health and Welfare, Ministry of Agriculture, Forestry and Fisheries, Ministry of International Trade and Industry, Ministry of Transport, Ministry of Posts and Telecommunications, Ministry of Labour, Ministry of Construction, and the Ministry of Home Affairs.

Science and Technology Agency (1997). Promotion of Strict Assessment of Research and Development. Tokyo: Science and Technology Agency; Science and Technology Policy Bureau; Planning and Evaluation Division; Office of Evaluation.
http://www.sta.go.jp/shokai/98eg/2_18.html

The strict assessment of government research and development is absolutely vital to stimulate and improve the efficiency of research and development activities, and to produce better results. In the Science and Technology Basic Plan, the implementation of strict research and development assessments is positioned as an extremely important policy. In order to facilitate the implementation of such assessments, the „General Guidelines Concerning the Future Direction of Concerning the Future Direction of Evaluation Common to Government Research and Development in General“ which act as guidelines for the implementation of assessments of government research and development were decided upon by the government in August 1997 after going through the process of review and written opinions in the Council for Science and Technology.
In addition to tackling the implementation of strict assessment in line with these guidelines, the Science and Technology Agency, by holding liaison councils with relevant ministries and agencies, is carrying out active support to ensure that strict assessments in line with these guidelines are carried out in the government as a whole.
(A) Implementation of Strict Assessments in the Science and Technology Agency: In response to the formulation these guidelines, steps are being taken for the implementation of strict assessments, such as promoting the development of a framework for the implementation of assessments in competent relevant research institutions and internal bureaus and departments.
(B) Promotion of Support for Initiatives Relating to Assessment of the Government as a Whole: In order to effectively promote assessment of research and development as the government as a whole, a liaison council comprising 15 relevant ministries and agencies was established, and it is promoting cooperation and support between relevant ministries and agencies concerning the implementation of assessments.
In addition to obtaining the understanding of government research and development and ensuring the transparency and impartiality of assessments, „The State of Assessment of Government Research and Development (Provisional Name)“ is published. This document presents in an easily understood manner all of the information of the government as a whole concerning assessment results implemented by the various ministries and agencies.

Tanaka, M. (1989). Japanese-Style Evaluation Systems for R&D Projects: The MITI Experience. Research Policy, 18, 361-378.

Tomizawa, Hiroyuki & Niwa, Fujio (1996). Evaluating overall national science and technology activity: General Indicator of Science and Technology (GIST) and its implications for S&T policy.

A method for overall evaluation of national S&T activities using a large number of quantitative indicators is proposed. Multivariate analysis was used to analyse the structure of a a set of 14 S&T indicators in the USA, Japan, Germany, France and the UK. The original indicators were integrated into a single indicator of S&T activities - the General Indicator of Science and Technology (GIST). The GIST is used to discuss the reforms of the Japanese Government in recent years in its S&T policies, thus showing the potential of the methodology as a tool for policy-makers.

Hochschulevaluation Japan

Alexander, Arthur J. (1999). University Research and Economic Growth in Japan. International Higher Education, Number 16, Summer, 10-11.

The business orientation of Japan's R&D was correctly identified in the past as the foundation of the country's technological strength. Now, that is a growing problem. In advanced countries, the linkages between basic research and the economy have intensified to such a degree that the practical orientation of much of Japan's scientific community and the acknowledged weaknesses of its basic research and university science may retard productivity growth in the future.

Arimoto, Akira (1998). The Changing Academic Evaluation System in Japan. Paper presented at the 11th Annual CHER Conference on the 10th Anniversary of the Consortium of Higher Education Researchers, „Higher Education Research – Achievements, Conditions and New Challenges“, 3-5 September 1998, Kassel, Germany.

Since 1991, self-reviews and evaluation procedures have been gradually established in universities and colleges around the country. In a national survey conducted in 1998, 83.7% of the 418 responding institutions replied that they had already conducted a self-review and evaluation more than one time.

Baba, Masateru (1996). The evaluation of the higher education system in Japan. In Robert Cowen (Ed.). World Yearbook of Education 1996: The Evaluation of Higher Education Systems (102-112). London: Kogan Page.

Darby, Michael R. & Zucker, Lynne G. (1996). Star Scientists, Institutions, and the Entry of Japanese Biotechnology Enterprises. Cambridge, Mass.: National Bureau of Economic Research, Working Paper 5795.

Hicks, Diana (1993). University-industry research links in Japan. Policy Sciences, 26, 361-395.

National Institute of Science and Technology Policy (1998). Survey and research into state of multi-disciplinary faculties in universities. Tokyo: National Institute of Science and Technology Policy (NISTEP), NISTEP Research Material & Data Report No. 53.

Yamamoto, Shinichi (1997). Research Evaluation and Universities in Japan: An Experience from the University of Tsukuba. In OECD (Ed.), The Evaluation of Scientific Research: Selected Experiences (91-97). Paris: OECD/GD(97)194.

12. Italien

Forschungsevaluation

Galante, E. & Sala, C. (1996). R&D evaluation at the Italian National Research Council: the agricultural sector. Scientometrics, 36(2), 207-222.

The principles and methodology of intra-mural and extra-mural research assessment developed at the Italian National Research Council are critically described.

MURST (1997). La Riforma del sistema ricerca Italia. Università Ricerca UR, anno VII, No. 3 (Special Issue).

A survey has been carried out by the Institute for Studies on Scientific Research and Documentation of the Ministry for University and Scientific and Technological Research to analyse institutions and activities related to research and innovation. This and related inquiries show the weak points of the Italian research system: (1.) insufficient human and financial resources, (2.) insufficient systemic approach, (3.) lack of systematic assessment, (4.) insufficient planning and weakness in the formulation of strategic programmes, (5.) inadequate evidence and diffision of results achieved by the research system.

Silvani, Alberto & Sirilli, Giorgio (1995). R&D evaluation in Italy: a science and technology policy view. Research Evaluation, 5(1), 69-77.

The situation in Italy with regard to evaluation is unsatisfactory: the limited experiences in carrying out evaluation, the inherent risks of its improper use, and the difficulty of reporting results in the presence of unclear objectives represent the major obstacles. However, progress on the adoption of evaluation procedures has come in recent years. In this paper the major evaluation studies are reviewed in the context of Italien science and technology policy. It is argued that the limited diffusion of evaluation in Italy is linked to cultural, organisational and institutional factors. The increasing demand by the public for accountability, and the growing internationalisation of R&D, accelerated by European integration, is expected to contribute to the spreadingof evaluation processes in the country.

Sirilli, Giorgio & Meliciani, Valentina (1994). Research evaluation at the National Research Council of Italy: a survey of decision-makers. Research Evaluation, 4(2), 75-88.

The paper investigates the evaluation of R&D at the National Research Council of Italy (CNR) on the basis of the results of interviews with the Presidents of National Advisory Committees and the Directors of the Mission-Oriented Projects. The results show that the evaluation process is based, to the extent to which it is performed, almost exclusively on scientific criteria, that there is a lack of standardised methodologies and only modest attention is paid to the evaluation of the social and economic impact of research activities. At the same time Presidents and Directors appear to be conscious of the Italien backwardness in this field and of the need to increase research evaluation in order to move towards a more efficient and effective use of resources.

Hochschulevaluation Italien

Boffo, Stefano & Moscati, Roberto (1998). Evaluation in the Italian Higher Education System: many tribes, many territories, ...many godfathers. European Journal of Education, 33(3), 349-360.

13. Spanien

Forschungsevaluation

Presmanes, B., Casado, J. & Guerrero, H. (1999). A Unified Framework for R&D Evaluation and Foresight. The Role of ANEP (The National Agency für Evaluation and Foresight) in Spain. In: Assessing Assessments - European Experiences (43-55). Proceedings of a conference organized by the Danish Institute for Studies in Research and Research Policy in cooperation with The European Consortium for Political Research. Aarhus: The Danish Institute for Studies in Research and Research Policy.

Sanz-Menéndez, Luis (1995). Research actors and the state: research evaluation and evaluation of science and technology policies in Spain. Research Evaluation, 5(1), 79-88.

The paper describes the development of research evaluation in Spain. It assumes that research evaluation, and R&D policy and programme evaluation are embedded in the development of an R&D system and are characterised by general Spanish policy-making. Research evaluation in a context of delegation and as a self-organising system for research actors guaranteed by the state, has been strongly developed in the last few years; R&D policy and programme evaluation is less institutionalised. The explanation is linked to the sequence of reforms of the R&D system and to the set up of the first Spanish science and technology policy.

14. Österreich

Forschungsevaluation

Österreichische Akademie der Wissenschaften (1998). Mittelfristiges Forschungsprogramm 1996-2000 - Evaluationsergebnisse I. Wien.

Die Broschüre gibt Auskunft über die Ergebnisse von sechs der bisher durchgeführten (Selbst-)Evaluationen von AW-Forschungseinrichungen (Festkörperphysik; Limnologie; Weltraumforschung, Astromonie, Atmosphärenphysik; Geschichte Österreichs und des Donauraums; Sozialwissenschaften; Asienforschung) sowie über Grundsätze, Abläufe, Arbeitsweise der Evaluationsgruppen, Evaluationsthemen, schriftliche Vorbereitung der Evaluation, Evaluatoren, öffentliche Programmdiskussion und Konsequenzen aus der Evaluation; über die Ergebnisse der verbleibenden sieben Evaluationen wird im Jahr 2000 in „Evaluationsergebnisse 2“ berichtet werden.

Österreichische Biochemische Gesellschaft (1995). Evaluation of Biochemical Sciences in Austria. Wien: Bundesministerium für Wissenschaft, Forschung und Kunst.

The evaluation, coordinated by the European Molecular Biology Organization (EMBO), deals with: (1) the scientific output of the institutes, (2) scientific contacts outside the institute, (3) cooperations with industry, (4) support for young scientists, (5) some aspects of teaching (as far as diploma and doctoral students are concerned), (6) cooperations of institutes in research, teaching, and services, (7) achievements and working load of institute members, and (8) strategies, plans and aims against the background of foreseeable developments in biochemistry.

Stampfer, M. (1997). Science and Technology Policy Evaluation in Austria: Struggling Towards a Higher Ranking on the Policy Agenda. In OECD (Ed.), Policy Evaluation in Innovation and Technology Towards Best Practices (Chapter 22). Paris: OECD. http://www.oecd.org/dsti/sti/stat-ana/prod/evaluation.htm

Steiner, Michael & Sturn, Dorothea (1995). Elements of evaluation of science and technology policy in Austria. Research Evaluation, 5(1), 98-108.

Within the last two decades there has been a remarkable intensification of activities in science and technology in Austria, and these changes have recently been accompanied by distinct efforts towards the development of evaluation processes. Evaluation mainly concentrates on Austria's technology policy, on technology support programmes, and on transfer and diffusion activities. There is a persistent lack of evaluation of research policy and university research. A new dimension of evaluation has been opened up by Austria's membership of the European Union, but as yet no standardised tools have been developed, and there is no consistent framework or process of feedback for the evaluation process. There are elements of evaluation within Austria, but as yet there is no system of evaluation.

Winter, Hannspeter (1992). Evaluation der physikalischen Forschung in Österreich - Erfahrungen und Empfehlungen. In Bundesministerium für Wissenschaft und Forschung (Hg.). Die Bewertung von Leistungen im Bereich von Lehre und Forschung (129-134). Wien.

Hochschulevaluation Österreich

Altrichter, Herbert, Schratz, Michael & Pechar, Hans (Hrsg.). (1997). Hochschulen auf dem Prüfstand. Was bringt Evaluation für die Entwicklung von Universitäten und Fachhochschulen? Innsbruck / Wien: Studien Verlag.

BOKU - Universität für Bodenkultur Wien (1998). Überlegungen zum Leistungsvertrag BOKU. Wien: BOKU. http://www.boku.ac.at/bdr/vrf/vrfbud~1.html

Bundesministerium für Wissenschaft und Forschung (1993). Evaluation im Hochschulwesen. Wien.

Die frühen neunziger Jahre sind eine Zeit tiefgreifender Veränderungen für das österreichische Hochschulsystem. Die Evaluation von Hochschulen hat einen neuen Stellenwert erhalten. Die politischen und rechtlichen Traditionen Österreichs, vor allem die Bildungsstrukturen des Landes, haben die Entwicklung solcher Verfahren bislang nicht gefördert. Wo von vornherein das Geschehen im Hochschulbereich durch detaillierte gesetzliche Regelungen determiniert ist, hält sich das Verlangen nach einer nachträglichen Überprüfung von Leistungen in Grenzen. Durch die Reform der Hochschulorganisation ändert sich diese Situation grundlegend. Evaluation ist seit den späten achtziger Jahren ein Schlüsselbegriff der hochschulpolitischen Diskussion geworden. Der Text verfolgt zwei Zielsetzungen: Zum einen will er dazu beitragen, die Diskussion über unterschiedliche Formen und Methoden der Evaluation auf eine breitere Basis zu stellen und auch die Erfahrungen anderer Länder miteinzubeziehen. Zum anderen wird die Möglichkeit zur Diskussion gestelt, auf der Ebene des Bundesministeriums für Wissenschaft und Forschung ein System von Leistungskennziffern zu entwickeln.

Bundesministerium für Wissenschaft und Forschung (Hg.). (1992). Die Bewertung von Leistungen im Bereich von Lehre und Forschung. Wien.

Angesichts der geplanten Universitätsreform, die weitreichende Dezentralisierungs- und Deregulierungsmaßnahmen vorsieht, bedarf es neuer Regelungsmechanismen zur Sicherung, Steigerung und Kontrolle universitärer Leistung. Evaluationen von Forschung und Lehre können in diesem Sinne als geeignetes Instrument angesehen werden, die Verteilung der verfügbaren Ressourcen zu steuern und qualitätsbezogene Veränderungen im universitären Lehr- und Forschungsbetrieb zu initiieren.

Bundschuh, E. (1999). Quality assessment and structural change in universities. In: European Commission & Austrian Advisory Board for Universities (Eds), Science and the Academic System in Transition: An International Expert Meeting on Evaluation (31-37). Budapest: Akadémiai Kiadó.

Campbell, David F. J. (1999). Evaluation universitärer Forschung. Entwicklungstrends und neue Strategiemuster für wissenschaftsbasierte Gesellschaften. SWS-Rundschau, 39(4), 363-383.

Die westlichen Demokratien gelten als wissenschaftsbasierte Gesellschaften, in denen der universitären Forschung eine wachsende Bedeutung zukommt. Die Evaluation der universitären Forschung trägt dabei zu deren Optimierung bei. Für die konzeptionelle Begründung von Evaluationen sind vier Argumente bedeutend: Evaluationen als (1) Instrumentarium für Feedback, (2) als Marktersatz, (3) als Instrumentarium für Rationalitätssteigerung und (4) Evaluationen als Nachweis für Ressourcenverwendung. Von diesen Argumenten abgeleitet wird im Artikel aufgezeigt, warum die ex post Evaluation universitärer Forschung von entscheidender Relevanz ist. Nach der Darstellung von ex post Evaluationsmodellen in Großbritannien und in den Niederlanden konzentriert sich der Beitrag abschließend auf Österreichs und Deutschlands Universitäten: Im Zentrum steht die Analyse von Strategiemustern für die flächendeckende Anwendung von Evaluationen.

Campbell, David F. J. (1999). New Challenges for the Evolution of Academic Research Systems: Four Hypotheses on the Evaluation of Academic Research. In: Susanne Bührer & Stefan Kuhlmann (Eds.), Evaluation of Science and Technology in the new Europe. Proceedings of an International Conference on 7 and 8 June 1999, Berlin (149-154). Bonn / Bruxelles: Federal Ministry of Education and Research / European Commission.

Campbell, David F. J. & Bernhard Felderer (1999). Empfehlungen zur Evaluation universitärer und außeruniversitärer Forschung in Österreich. Wien: Institut für Höhere Studien, Reihe Politikwissenschaft / Policial Science Series No. 66.

Die Herausforderung besteht darin, für Österreichs universitäre Forschung ein Evaluations-Gesamtmodell zu designen, das sich flächendeckend und systematisch anwenden läßt. Der Entwurf sieht einen dualen Evaluationsmodus vor, der zwei Evaluationsansätze gleichberechtigt miteinander kombiniert: einerseits ein Monitoring sowie andererseits eine externe ex post Evaluation von universitärer Forschung. In Abhängigkeit von den Evaluationsergebnissen und auf Basis einer Forschungspunkteformel können den Universitätsinstituten Forschungspunkte zugeteilt werden. Davon lassen sich wiederum systematische Konsequenzen für Ressourcenallokationen ableiten. Neben der Evaluation universitärer Forschung wird ein Evaluations-Gesamtmodell für die außeruniversitäre Forschung entworfen, das Schnittstellen zum universitären Evaluationsmodell aufweist. Ferner werden Empfehlungen für die Forschungsfinanzierung und für die Forschungsförderung des FWF (Fonds zur Förderung der wissenschaftlichen Forschung) zur Diskussion gestellt.

Campbell, David F. J. & Bernhard Felderer (1997). Evaluating Academic Research in Germany. Patterns of Policies. Wien: Institut für Höhere Studien, Reihe Politikwissenschaft / Policial Science Series No. 48.
Available via E-mail: library@ihs.ac.at

Felderer, Bernhard @ David F. J. Campbell (1999). Wie kann oder wie soll Österreichs akademische Forschung evaluiert werden? Empfehlungen zur Evaluation universitärer und außeruniversitärer Forschung in Österreich. Endbericht der Studie „Die Evaluation der akademischen Forschung im internationalen Vergleich: Strukturen, Trends und Modelle“ im Auftrag des Bundesministeriums für Wissenschaft und Verkehr. Wien: Institut für Höhere Studien.

Im Rahmen der Studie wurde einerseits ein bibliometrischer Publikationsvergleich der OECD-Nationen vorgenommen und andererseits ein Policy-Vergleich von akademischer Forschungsevaluation in fünf westeuropäischen Nationen: Deutschland, Großbritannien, die Niederlande, Finnland und die Schweiz. Der Schwerpunkt der Empfehlungen konzentriert sich auf die Universitäten, da in Österreich gegenwärtig fast vier Fünftel der akademischen Forschungsaufwendungen im Hochschulsektor durchgeführt werden. Die Empfehlungen beziehen sich auf vier Themenbereiche: (1) Empfehlungen zur Finanzierung der Forschung und der akademischen Forschung; (2) Empfehlungen zur Evaluation der universitären Forschung; (3) Empfehlungen zur Evaluation der außeruniversitären Forschung; und (4) Empfehlungen zur Forschungsförderung des FWF (Fonds zur Förderung der wissenschaftlichen Forschung).

Hackl, Elsa, Schwab-Matkovits, Ingrid, Linhart, Markus & Verdonk, Desireé (1997). In: Herbert Altrichter, Michael Schratz & Hans Pechar (Hrsg.), Hochschulen auf dem Prüfstand (144-154). Innsbruck: Studien-Verlag.

In Österreich wurde 1993 ein Fachhochschul-Sektor etabliert. Im Fachhochschulbereich wurde auf eine detaillierte rechtliche Vorgabe der Organisation, des Studienangebotes, des Dienst- und Besoldungsrechtes verzichtet. Das im Vergleich zum übrigen Bildungswesen geringe Maß an Input-Normierung macht Kontrollen der Effektivität und Effizienz von Tätigkeiten und Maßnahmen unabdingbar.

Massimiani, Roberto (1998). Aufgabenorientierte Budgetierung an Universitäten. Wien: Bundesministerium für Wissenschaft und Verkehr.

Mit dem Universitäts-Organisationsgesetz 1993 wurde die Ressourcenbewirtschaftung der Universitäten dezentralisiert und folgerichtig eine dezentrale Ressourcenplanung und eine neue Universitätsbudgetierung eingeführt. Die Universitäten erhalten nunmehr - der Baubereich ausgenommen - ein jährliches Gesamtbudget für Personal-, Sach- und Betriebsausgaben. Die jährliche Budgetzuweisung aus dem Bundeshaushalt ist abhängig von der Entwicklung des Leistungsangebotes der einzelnen Universitäten, berücksichtigt die budgetären Auswirkungen von universitätsrelevanten Gesetzen und sieht außerdem eine Sonderregelung für Großinvestitionen vor. Die Ressourcenplanung ist für die Lehre output-orientiert und für den Forschungsbereich input-orientiert konzipiert. Die Budgetzuweisung des Bundes folgt nicht dem Modell einer leistungsorientierten Budgetierung, da es äußerst unwahrscheinlich ist, daß mit der jährlichen Budgetzuweisung des Bundes der hochkomplexe Leistungsprozeß der Universitäten qualitätssteuernd beeinflußt werden kann.

15. Australien:

Forschungsevaluation

Australian Bureau of Industry Economics (1996). Australian Science: Performance from Published Papers. Canberra: Australian Government Publishing Service (Report 96/3).

Bourke, Paul (1997). Evaluating University Research: The British Research Assessment Exercise and Australian Practice. Canberra: National Board of Employment, Education and Training; Australian Science Council; Higher Education Council. Commissioned Report No. 56.

Bourke, P. & Butler, L. (1996). Publication types, citation rates and evaluation. Scientometrics, 37(3), 473-494.

Linke, Russell D. (1995). Evaluation of Research Performance: A Review of Selected Input and Output Characteristics. Canberra: Australian Government Publishing Service. (ISBN 0-644-35927-7)

Phillimore, John & Marinova, Dora (1993). Western Australia: Leading the Way to a „Clever Country“ - Trends in Science and Technology Indicators. Western Australia: Murdoch University, Institute for Science and Technology Policy.

In this paper, a variety of data sources are examined to provide a set of indicators of the current state of science and technology in Western Australia.

Turpin, Tim (1997). CRCs and Transdisciplinary Research: What are the Implications for Science? Prometheus, 15(2), 253-265.

A number of authors have recently proposed a future for science, where the traditional academic mode of knowledge production, primarily organised on disciplinary line, is largely replaced by a different mode of knowledge production that is more transient in its organisational forms. If correct, the new mode of knowledge production has implications for the research cultures of universities, government research institutes, or industrial laboratories. But in particular, the trend has implications for research arrangements, such as Cooperative Research Centres (CRCs).

Wood, Fiona Q. (1997). Peer Review Deconstructed. In: Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Ed.), The Future of the Peer Review System (53-87). The Hague: NWO.

The paper describes the process of peer review as it relates to the award of research grants. This chapter is followed by an overview of some of the major criticisms of the process. It then discusses some of the improvements and alternatives that have been suggested or put in place to offset these criticisms. Some reference is made to research grant funding issues in the Australian context.

Wood, Fiona Q. (1997). The Peer Review Process. Canberra: Australian Research Council (Commissioned Report Number 54).

Hochschulevaluation Australien

Harman, Grant (1998). Quality Assurance Mechanisms and Their Use as Policy Instruments: major international approaches and the Australian experiences since 1993. European Journal of Education, 33(3), 331-348.

Murphy, Penelope S. (1995). Benchmarking Academic Research Output in Australia. Assessment & Evaluation in Higher Education, 20(1), 45-57.

The paper presents benchmarks for publications productivity for Australian academic researchers based on an empirical study of all reported publications output from the Australian university system in 1991. Applicability within Australian higher education of citation impact indices as an indicator of research quality is discussed.

Sheehan, Barry (1996). The evaluation of higher education system in Australia. In Robert Cowen (Ed.). World Yearbook of Education 1996: The Evaluation of Higher Education Systems (14-33). London: Kogan Page.

Turpin, Tim, Garrett-Jones, Sam, Rankin, Nicole, Aylward, David & Johnston, Ron (1996). Patterns of Research Activity in Australian Universities. Phase One: Final Report (Chapter 5: Research Output across Universities, S. 73-82). Canberra: National Board of Employment, Education and Training, Australian Research Council. Commissioned Report No. 47. ISBN 0-644-39669-9

Turpin, Tim, Aylward, D., Garrett-Jones, S., Speak, G., Grigg, L. & Johnston, R. (1999).

16. Europäische Kommission

Forschungsevaluation

Aguado-Monsonet, Miquel A. (1998). Use of bibliometrics as a technology watch technique. Application to the analysis of the recent developments of the photocatalysis. Sevilla: Institute for Prospective Technological Studies, European Commission, Joint Research Centre. EUR-18131-EN
http://www3.jrc.es/scripts/dbWeb/dbwebc.dll/ipts?linkxresults/obj/ipts/col/ID/dat/21
http://www3.jrc.es/scripts/dbWeb/dbwebc.dll/ipts?getxresults
http://www.jrc.es/pages/ourrole/products.html

The aim of this report is to demonstrate the usefulness of bibliometrics as a technology watch technique. It is known that a wide range of bibliometric tools are available. But within this report we have concentrated the effort in those bibliometrics tools that can serve to carry out a technology watch analysis. Bibliometrics is as a powerful technology watch technique that can provide decision makers with a quick picture of the state of the art of a specific domain of the science or technology.

Aguilar, Alfredo, Ingemansson, Torbjörn, Hogan, Stéphane & Magnien, Etienne (1998). Peer review evaluation of proposals in the biotechnology programme of the European Union. Research Evaluation, 7(3), 141-146.

The first part of the paper describes the evaluation procedure or peer review undertaken by the European Union biotechnology programme. The second part assesses the evaluation procedure, through an analysis of the responses to a questionnaire sent to all experts who have participated in the evaluation of proposals for the current biotechnology programme of the Fourth Framework Programme (1994-98). The overall feedback from evaluators is of general satisfaction. Panel discussions have sometimes led to policy-making, for instance, in the 'clustering' of research initiatives with a potential for synergy and wider spin-offs. Evaluation reports have triggered daring initiatives to cross-link new technology trends with the needs of a diverse community of users.

Airaghi, Angelo, Baptista, José Viana, Busch, Niels E., Georghiou, Luke, Kuhlmann, Stefan, Ledoux, Marc J. & Raan, Anthony F. J. van (1999). Options and Limits for Assessing the Socio-Economic Impact of European RTD Programmes. Brussels: Commission of the European Communities, ETAN Expert Working Group, January 1999.

Bach, Laurent & Georghiou, Luke (1998). The Nature and Scope of RTD Impact Measurement. A discussion paper for the International Workshop on „Measurement of RTD Results/Impact“, Brussels, 28-29 May 1998. Strasbourg: L. Pasteur University of Strasbourg, BETA, and Manchester: University of Manchester, PREST.

Barré, Rémi (1997). The European perspective on S&T indicators. Scientometrics, 38(1), 57-70.

Bührer, Susanne & Kuhlmann, Stefan (1999). Evaluation of Science and Technology in the new Europe. Proceedings of an International Conference on 7 and 8 June 1999, Berlin. Bonn / Bruxelles: Federal Ministry of Education and Research / European Commission.

Callon, Michel, Larédo, Philippe & Mustard, Philippe (1995). La gestion stratégique de la recherche et de la technologie. Paris: Economica (ISBN 2-7178-2853-2).

The book contains contributions of 28 experts from England, France, Germany, the Netherlands, Finland and Switzerland. It offers an excellent survey of the state of the art in research evaluation, as reflected in the accumulated experience within the European Union over the last 15 years.

Callon, Michel, Larédo, Philippe & Mustard, Philippe (1997). The Strategic Management of Research and Technology - Evaluation of Programmes. Paris: Economica International.

Cunningham, P. (1997). The evaluation of European Programmes and the future of scientometrics. Scientometrics, 38(1), 71-85.

The paper presents the results of an examination of a selection of published European evaluations. The incidence of quantitative and scientometric approaches has been reviewed and an assessment made of their contributory role in each evaluation.

Dumont, Y., Durieux, L., Karatzas, I., O’Sullivan, L., Teuber, H., Stroud, G. & Fayl, G. (1999). EU RTD Programmes Impact/Results Assessment: A Manifold Task in a Complex System. In: Susanne Bührer & Stefan Kuhlmann (Eds.), Evaluation of Science and Technology in the new Europe. Proceedings of an International Conference on 7 and 8 June 1999, Berlin (59-69). Bonn / Bruxelles: Federal Ministry of Education and Research / European Commission.

Durieux, L. & Fayl, G. (1997). The Scheme Used for Evaluating the European Research and Technological Development Programmes. In OECD (Ed.), Policy Evaluation in Innovation and Technology Towards Best Practices (Chapter 20). Paris: OECD. http://www.oecd.org/dsti/sti/stat-ana/prod/evaluation.htm

European Commission (1997). Second European Report on S&T Indicators 1997 (Chapter 2: S&T performance of Europe). Luxembourg: Office for Official Publications of the European Communities. EUR 17639 EN.

European Commission (1997). Five-year assessment of the European Community RTD framework programmes. Luxembourg: Office for Official Publications of the European Communities. EUR 17644 EN (ISBN 92-828-0248-5)

European Commission: Five-year assessments (since 1995)
- Industrial Materials and Technology (EUR 17587)
- Standards, Measurements and Testing (EUR 17588)
- Environment & Climate (EUR 17589)
- Marine Science & Technologies (EUR 17590)
- Biotechnology (EUR 17591)
- Biomedicine & Health (EUR 17592)
- Agriculture, Fisheries, Forestry & Agro-Industry (EUR 17593)
- Non-nuclear Energy (EUR 17594)
- Transport (EUR 17595)
- Targeted Socio-Economic Research (EUR 17596)
- Human Capital & Mobility (EUR 17598)
- Nuclear Fission Safety (EUR 17599)
- Dissemination & Optimization of Results (EUR 17600)
- Information Technologies (EUR 17601)
- Advanced Communication Technologies (EUR 17602)
- Telematics Application (EUR 17603)

European Commission & Austrian Advisory Board for Universities (1999). Science and the Academic System in Transition: An International Expert Meeting on Evaluation. Budapest: Akadémiai Kiadó. (ISBN 963-05-7673-2)

The volume provides an overview of theories, methods and practical experiences gained in evaluation in diverse fields. Both the instruments of evaluation and forms of their application, for instance in academic contexts, in politics and in R&D, are placed in focus. The discussion is situated within the framework provided by the different European experiences gained in the theory and practice of evaluation, as well as by the role of the European Union in the politics of science and technology.

Fayl, Gilbert, Dumont, Yves, Durieux, Luc, Karatzas, Isidoros & O'Sullivan, Liam (1998). Evaluation of research and technological development programmes: a tool for policy design. Research Evaluation, 7(2), 93-97.

In 1994, the European Commission introduced a new scheme for the evaluation of its multi-annual research and technological development (RTD) programmes. The scheme combines two activities both involving independent experts: for each RTD programme, a continuous monitoring and a five-year assessment. The monitoring provides a major input to the assessment, which itself combines an ex post evaluation of the previous programme, a mid-tern appraisal of the ongoing one, and recommendations for future orientation. The recent five-year assessment (which resulted in more than 25 reports and involved some 200 European experts) provided input to the preparation of the Fifth RTD Framework Programme of the European Community (1998-2002). The new scheme has shown its ability to provide early feedback from evaluation into policy formulation.

Gabolde, Jean (1998). New challenges for indicators in science and technology policy-making: A European view. Research Evaluation, 7(2), 99-104.

Directorate-General XII of the European Union has always supported the integration of different kinds of S&T indicators. This resulted in the publication of the first report of S&T indicators in 1994 and the second one in 1997. Despite major efforts in this second report to bring together all different kinds of S&T input and output indicators at the European, national, regional and even enterprise level, there is still a large gap between the needs of the policy-makers and the existing indicators. Why? There is a well recognised evolution from a 'linear' to an 'integrated' science and innovation system. This calls for new indicators of connections, interactivity, linkage, in short, for

a new family of 'systemic indicators'. The paper also covers how this evolution is taken into account by key actions at the European level.

Georghiou, Luke (1999). Accountability to the Public and Value-for-Money. In: Susanne Bührer & Stefan Kuhlmann (Eds.), Evaluation of Science and Technology in the new Europe. Proceedings of an International Conference on 7 and 8 June 1999, Berlin (133-135). Bonn / Bruxelles: Federal Ministry of Education and Research / European Commission.

Georghiou, Luke (1999). Socio-economic effects of collaborative R&D - European experiences. Journal of Technology Transfer, 24(1), 69-79.

Georghiou, L., Dale, A. & Cameron, H. (Eds.). (1995). National systems for evaluation of R&D in the European Union (Special Issue). Research Evaluation, 5(1), 1-108.

Guzzetti, Luca (1995). A brief history of European Union Research Policy. Luxembourg: Office for Official Publications of the European Communities. (ISBN 92-827-5353-0)

Karatzas, Isidoros, O'Sullivan, Liam, Fayl, Gilbert, Durieux, Luc, & Dumont, Yves (1998). Evaluating the European Commission RTD Programmes: Responding to the Needs of a Changing World. In Vaclav Paces, Ladislav Pivec & Albert H. Teich (eds.), Science Evaluatiion and Its Management (31-35). Ohmsha: IOS Press, NATO Science Series: Science & Technolgy Policy, vol. 28.

Evaluation has been a legislative requirement for European Commission Research and Technology Development (RTD) programmes since the early 1980's. The changing Science and Technology environment and the increased pressure for timely, independent evaluation led to the adoption of the current evaluation scheme which has been in force for the last three years. The new scheme is based on annual monitoring carried out with the assistance of independent experts and a five-year assessment conducted by independent experts. The five-year assessment combines an ex-post evaluation of the previous programme, a mid-term appraisal of the on-going one and recommendations for future orientation(s). The new scheme is a tool for programme management and provides timely and independent feed-back into policy formulation.

Kuhlmann, Stefan, Boekholt, Patries, Georghiou, Luke, Guy, Ken, Héraud, Jean-Alain, Laredo, Philippe, Lemola, Tarmo, Loveridge, Denis, Luukkonen, Terttu, Polt, Wolfgang, Rip, Arie, Sanz-Menendez, Luis & Smits, Ruud (1999). Improving Distributed Intelligence in Complex Innovation Systems (Chapter 2. 2.: Innovation Policy Evaluation, S. 31-40). Final report to the Advanced Science & Technology Policy Planning Network (ASTPP). Brussels: EU/TSER.

The European evaluation culture has a broad range of conceptual and methodological experiences at its disposal. Methods of various types have been developed and utilised to determine attained or attainable effects; the most important are: peer reviews, before / after comparisons, control or comparison group approaches, a variety of quantitative and qualitative analyses, etc. These conceptios can be carried out individually or in combination with various data and indicators (financial expenditure on R&D, patents, economic, social, technical indicators, publications, citations, etc.), data collection methods (existing statistics, questionnaires, interviews, case studies, panels, etc.), data analysis methods (econometric models, cost/benefit analyses, other statistical methods, technometrics, bibliometrics, peer review, etc.). All the procedures have different strengths and weaknesses, which makes using a combination of methods advisable.

Krull, W., Sensi, D. & Sotiriou, D. (1991). Evaluation of R&D - Current Practice and Guidelines (Synthesis Report). Luxembourg: Office for Official Publications of the European Communities (catalogue no. CD-NA-13336-EN-C; ISBN 92-826-2232-0).

Kyriakou, D. (1995). Macroeconomic aspects of S/T programme evaluation. Scientometrics, 34(3), 451-459.

Larédo, Philippe (1995). Structural effects of EC RT&D programmes. Scientometrics, 34(3), 473-487.

Luukkonen, Terttu (1998). The difficulties in assessing the impact of EU framework programmes. Research Policy, 27, 599-610.

Ormala, E. (1994). Impact Assessment: European Experience of Qualitative Methods and Practices. Evaluation Review, 18, 41-51.

Paterson, George (1999). Measuring the stocks and flows of human resources in science and Technology. Research Evaluation, 8(2), 91-97.

As the 21st century approaches, and more and more countries move towards the knowledge-based economy, the need for S&T indicators becomes more pressing. One key area yet to be fully developed is indicators measuring the human resources in S&T. The paper presents the latest work in Eurostat to develop a series of indicators, using existing international data, capable of comparing human resource potential throughout the EU.

Peterson, John & Sharp, Margaret (1999). Technology Policy in the European Union. St. Martins Press. ISBN: 0312216416

Removille, J. & Clarysse, B. (1999). Intra-European scientific co-operation: measuring poliy impact. Research Evaluation, 8(2), 99-109.

European Union (EU) funded co-operation within the Community programme and the other public S&T co-operation taking place outside the specific remit of the EU, has grown from 4% (of government R&D expenditure) in 1985 to 16% in 1995. An attempt is made to gain insight into how different public research programmes influence the structure of S&T co-operation. It is found that while the pre-competitive research networks have become quite international, market-driven networks (e. g., Eureka networks) remain culturally anchored.

Rinaldini, C. (1995). Experience on research evaluation at the Joint Research Centre of the European Commission. Scientometrics, 34(3), 519-525.

Whitney, Gretchen (1993). Patterns of authorship in major bibliographic databases: The European region. Scientometrics, 26(2), 275-292.

European authorship trends in fifteen major scientific and technical bibliographic databases on the DIALOG information system are examined for works published between 1970 and 1990. Overall, with the exception of MEDLINE, BIOSIS, and INSPEC coverage of the works of European authors has been declining over the past twenty years, and particularly so in the last five.

Hochschulevaluation Europäische Kommission

Boffo, Stefano, Chave, Daniel, Kaukonen, Erkki & Opdal, Liv Randi (1999). The Evaluation of Research in European Universities, 34(3), 325-334.

The EVALUE study shows that the European universities differ mostly in their profile and that the strength and orientation of their research vary greatly. This comes out in our classification of three ideal types of universities: the comprehensive, the applied/professional and the regional universities. On the basis of the EVALUE-study, some general trends in research evaluation are outlined: (1.) The change of emphsis from an internal, traditional evaluation to a more external evaluation. (2.) The change from an individual, traditional evaluation to a more collective one. (3.) The change from qualitative to quantitative criteria. (4.) Most of the evalluations stress their

analyses of the quality of the research and not only its quantity. (5.) There is also a trend to refer to an international standard of research evaluation, and (6.) a trend towards a greater financial impact of the evaluation.

The Centre for Quality Assurance and Evaluation of Higher Education (Denmark) in cooperation with Comité National d'Evaluation (France) (1998), Evaluation of European Higher Education - A Status Report. Prepared for the European Commission, DG XXII. Brussels.

Over the last decade the focus on evaluation as a steering mechanism as well as a tool of improvement has increased remarkably. The aim of the report is to present an overview (as it was in the spring of 1998) of the state of the art of evalution in the 15 European member states and the two EEA-countries, Norway and Iceland.

Committee for Higher Education in the European Community (CHEEC), with the collaboration of the Center for Higher Education Policy Studies (CHEPS) of the University of Twente (1993). Quality management and quality assurance in European higher education: methods and mechanisms. Luxembourg: Office for Official Publications of the European Communities. (ISBN 92-826-6391-4)

The report presents the methods and mechanisms used in European Community Member States and the EFTA countries to manage the quality of higher education. These include both traditional procedures and new approaches introduced from the early 1980s onwards. The new methods and procedures of quality management have so far been implemented most widely in France, the Netherlands and the United Kingdom. Common elements of the new methods of quality management are: (1) the meta-level role of managing agent(s), (2) the meachanism of self-evaluation, (3) the mechanism of peer review and site visits, (4) the reporting of the results, and (5) the possible relationship between quality review outcomes and funding of higher education activities.

Dubois, Pierre et al. (1998). EVALUE: Evaluation and Self-Evaluation of Universities in Europe (Project Report). Luxembourg: Office for Official Publications of the European Communities.

The project, funded under the Targeted Socio-Economic Research (TSER) Programme of the Fourth Framework Programme, brought together 12 research teams in a collaborative endeavour. The research carried out analyses the procedures of self-evaluation and the knowledge of its impact on the performance of universities in different European countries. The project developed an assessment procedure for higher education institutions, teachers and administrative staff which is pluralistic, context-sensitive and dynamic.

Scheele, Jacob P., Maassen, Peter A. M. & Westerheijden, Don F. (Eds.). (1998). To be continued... Follow-up of Quality Assurance in Higher Education. Maarssen: Elsevier/De Tijdstroom.

The book reflects the various approaches in countries of the European Union and the discussions about methods and responsibilities in the evaluation of follow-up of outcomes of quality-assurance.

Senker, J. M. (Ed.). (1999). European Comparison of Public Research Systems. Brussels: EC/TSER.
http://www.sussex.ac.uk/spru/psr/psr.html

The European Comparison of Public Research Systems project was funded by the EC's TSER programme. It began in March 1997, involved ten partners and was led by SPRU. It compared the changing organisation and structure of public sector research (PSR) in 12 European countries, and developed a methodology to examine how national policies affect researchers at bench level. The results are summarised in a downloadable report. They reflect growing convergence among the

previously rather distinctive functions of various sectors of PSR, a growth in the importance of university research and a casualisation of Europe's research labour force. In every country there has been increasing emphasis on the promotion of economic growth, innovation and technology transfer. Research collaboration of all types is increasing. There are also common trends in management practices across countries such as evaluation, and mechanisms for strategic planning in various forms. The differing characteristics of each national PSR system, however, produce diverse results from these practices. Each country can learn from others; but what it learns must be tailored to the specific national PSR system.

Trinczek, Rainer & West, Anne (1999). Using Statistics and Indicators to Evaluate Universities in Europe: aims, fields, problems and recommendations. European Journal of Education, 34(3), 343-356.

At present, it is often not so much the number or kinds of statistics/indicators produced in the different countries of Europe that cause the major problems, but a lack of comparability and user-friendliness.

17. European Science Foundation

European Science Foundation (1999). An ABC of Medical Research Funding. Strasbourg: ESF.

A short easy-to-follow ESF Guide to appraisal and administration of research grant applications with tips on how to write and present them.

18. OECD

Forschungsevaluation

Aubert, J.-E. (1997). On the OECD Experience of Country Reviews. In OECD (Ed.), Policy Evaluation in Innovation and Technology Towards Best Practices (Chapter 21). Paris: OECD.
http://www.oecd.org/dsti/sti/stat-ana/prod/evaluation.htm

OECD (1997). The Evaluation of Scientific Research: selected experiences. Paris: OECD, Committee for Scientific and Technological Policy. Document OECD/GD(97)194.
http://www.oecd.org/dsti/sti/s_t/scs/prod/e_97-194.htm

Research evaluation has emerged as a „rapid growth industry“. In most OECD countries there is an increasing emphasis on accountability, as well as on the effectiveness and efficiency of government-supported research. A series of concrete experiences, presented at a workshop held at OECD in April 1997, illustrate various approaches to research evaluation at both the country and institutional levels.
Research evaluation efforts can be focused on different entities differentiated by increasing levels of size and complexity. At the first level, evaluation can focus on the work of individual researchers. Second, it can concern larger research groups, laboratories and institutions such as universities. Third, evaluation can focus on an entire scientific discipline. Fourth, it can concern government programmes and funding agencies. Finally, evaluation methodologies can be applied to a country’s entire research base.

The evaluation of researchers as individuals is illustrated by experience of the French „Centre National de la Recherche Scientifique“ (CNRS) where evaluation is performed as a basic instrument for personnel management and promotion. The Swedish Natural Science Research Council's (NFR) experience of international evaluation concerns both researchers and disciplines, while offering the same token views on government support to specific people and projects. The Japanese experience shows how evaluation practices are being developed in universities as stimulated by recent government guidelines which aim to raise the level and significance of basic research conducted in the country's universities. The German experience presents the methods used for rationalising the network of some 80 government R&D institutes, known as the „Blue List“ institutes. The US National Science Foundation's (NSF) experience deals with evaluation of both university research and education performance. Finally, three overviews of evaluation practices covering an entire research system at the country level were presented by the experiences of Finland, the Netherlands and the United Kingdom.

Whatever the subjects or level of the entities evaluated, and whatever the evaluation "culture" of the concerned country, research evaluation depends on two basic, complementary approaches: the use of quantitative indicators, such as bibliometrics, on the one hand, and the use of more qualitative peer judgements, on the other. Some research cultures are more concerned with the need for detailed output measures than others and pay particular attention to quantifiable indicators. In contrast, other cultures tend to limit the use of quantitative indicators.

However, evaluation should not be considered as an end in itself. Rather, it should be developed and used more as a pointer to key policy issues and essential questions that need to be addressed. Research evaluation becomes useful to the extent that it helps in clarifying policy debates and moves decision-making processes forward on more rational and quantifiable grounds that improve the understanding of all partners involved in such decision making.

OECD (1997). Policy Evaluation in Innovation and Technology: Towards Best Practices. Paris: OECD.
http://www.oecd.org/dsti/sti/stat-ana/prod/evaluation.htm

Evaluation of government programmes and policies is an issue of increasing interest in OECD countries. It is driven by tight budgets, a greater focus on accountability and transparency in policy, and the desire to minimie distortions arising from government actions while maximising their impact. In the innovation and technology area, policies aim to improve the capacity of firms to innovate and use new technologies, thus contributing to higher productivity and growth, and to the creation of more and better jobs. Given the growing importance of knowledge-based economic activities, it is crucial to be able to identify how the maximum leverage of these policy initiatives can be obtained. Evaluation is thus central to „best practice“ formulation in thus area.

This report brings together presentations by a range of international researchers and policy makers at an OECD Conference held on 26-27 June 1997. The contributions have been reorganised for publication in order to better reflect methodological issues as well as very different country experiences. They provide valuable insight into the evaluation of innovation and technology practices in OECD countries, focusing in particular on the quantitativie and qualitative tools used and the institutional set-up within which evaluation exercises take place.

OECD (1995). Cost/Benefit Analysis of Large S&T Projects: Some Methodological Issues. Paris: OCDE/GD(95)57.
http://www.oecd.org/dsti/sti/s_t/ms/prod/e_95-57.pdf

Schmoch, Ulrich (1999). Impact of international patent applications on patent indicators. Research Evaluation, 8(2), 119-131.

International patent applications, according to the Patent Co-operation Treaty, enjoy increasing popularity, as they provide various advantages for the applicants. The enormous growth of this type of application path has a relevant impact on patent statistics and therefore on patent indicators. The most problematic aspect is the delayed transfer of international applications to the national or regional office of destination, leading to an underestimation of recent application numbers. The article suggests a projection method compensating this effect in a reliable way.

Schmoch, Ulrich (1997). Indicators and the relationship between science and technology. Scientometrics, 38(1), 103-116.

The relationship between science and technology is an important issue, as science-based technologies play a key role in modern economies. The exploration of the science-technology interface can be effectively supported by quantitative indicators, in particular patents of scientific institutions, publications of industrial enterprises, and scientific references in patent search reports. The most promising appraoch is the parallel observation of patents and publications in order to analyse the dynamics of the interaction of science and technology and the professional move of academic and industrial researchers between institutions.

Hochschulevaluation OECD

Brennan, John & Shah, Tarla (1999). Institutional Experiences of Quality Assessment in Higher Education. Paris: OECD-CERI, IMHE Programme. http://www.oecd.org/els/edu/imhe/instexp.htm

In 1994, the OECD Programme on Institutional Management in Higher Education (IMHE) launched a new project entitled: „Quality Management, Quality Assessment and the Decision-Making Process“. The objectives were: (a) to clarify the purposes, methods and intended outcomes of different national systems of quality assessment, and (b) to investigate their impact on institutional management and decision-making. The project involved two main phases. Firstly, a conceptualisation and review of national systems of quality assessment in terms of their purposes and contexts was undertaken in 1995. Secondly, a series of institutional case studies was undertaken during 1996 on the impact of quality assessment on institutional management and decision-making.

Kells, H. R. (Ed.). (1990). The Development of Performance Indicators for Higher Education: A Compendium for Eleven Countries. Paris: OECD, Programme on Institutional Management in Higher Education (IMHE).

The report examines the development and implementation of performance indicators of higher education, through presentation of position statements on 11 countries: Australia, Austria, Canada, Denmark, Finland, France, Greece, Netherlands, Norway, Sweden, and United Kingdom. General conclusions are summarized in an introductory section and include the following observations: (1) there has been substantial development in the area of performance indicators; (2) developments clearly reflect the importance of national and cultural settings, and the political agendas of governments figure strongly in performance indicator development; (3) the emerging relationship between performance indicators and funding mechanisms is of importance; (4) institutional self-regulation within well-defined institutional and systemic goals is desirable; (5) some governments are requesting that institutions prepare 3-5 year development plans; (6) there is concern as to the adequacy of the management information systems needed to sustain performance indicators; (7) while there is some satisfaction with progress made on indicators relating to research, much work remains in the domains of teaching, finance, and public service; and (8) of greatest concern is the problem of publication of indicators on a comparative basis across institutions.

Kogan, Maurice (Ed.). (1993). Evaluating Higher Education (second impression). London: Jessica Kingsley Publishers (Higher Education Policy Series 6).

The articles in this book display the wide range of possible approaches to the evaluation of higher education that can be found at different levels in different systems. The experiences described here are drawn from many countries represented in the Organisation for Economic Cooperation and Development (OECD) but raise issues which concern higher education in all countries where questions of quality control and of accountability have become prominent. The book contains

chapters on (1) Approaches and Techniques of Evaluation, (2) Evaluating Institutions, (3) Evaluating Faculty, Courses and Departments, and (4) Evaluation of Research.

Massaro, Vin (1997). Learning from audit? Preliminary impressions from a survey of OECD countries. In Högskoleverket (Ed.). Quality assurance as support for processes of innovation - The Swedish model in cooparative perspective (9-38). Stockholm: Högskoleverket (The National Agency for Higher Education), Högskoleverket Studies 1997:1 S.

The author reports on the findings of a study he is conducting for the OECD and thereby gives an overview of the various models chosen by different countries, what they focos on and how they organise their quality assurance.

OECD (Ed.). (1999). Quality and Internationizing in Higher Education. Paris: OECD. ISBN 92-64-17049-9

This volume includes a general discussion of quality assurance issues in an international context, and case studies of specific institutions. The focus of the study is on the Internal Quality Review Process (IQRP), an effort to harmonize issues relating to quality assurance among academic institutions in different countries.

Shah, Tarla (1997). Quality Management, Quality Assessment and the Decision-Making Process: The IMHE Project on Institutional Impact. In John Brennan, Peter de Vries & Ruth Williams (Eds.), Standards and Quality in Higher Education (205-215). London: Jessica Kingsley Publishers.

The aims of the project are to examine both the declared purposes of different national quality agencies and the impact of the agencies, particularly on management and decision-making, as experienced by higher education institutions. A part of the project, over 40 institutions are conducting case studies of their own experiences of quality assessment. The interests of the IMHE (Institutional Management in Higher Education.) project are in the ways in which quality assessment enters into broader relationships between higher education and the state and into internal relationships within institutions, effecting in both cases a shift in the balance of power and the locus of decision-making.

OECD/IMHE-Internet-Adresse: http://www.oecd.org/els/edu/imhe/index.htm
List of publications: http://www.oecd.org/els/edu/index.htm and http://www.oecd.org/els/publicatio/els_publ.htm

19. NATO

Forschungsevaluation

Paces, Václav, Pivec, Ladislav & Teich, Albert H. (1999). Science Evaluation and its Management. Ohmsha: IOS Press, NATO Science Series: Science & Technolgy Policy, vol. 28 (ISBN:: 90 5199 438 9).

Evaluation of scientific research, particularly that research which is supported by government funds, is matter of growing concern in virtually every nation. No longer is it sufficient to expect that the value of investments in research will be judged in long-term, historical perspective. Resources are inevitably scare and policy-makers are seeking ways to assure that these resources are used in the most effective way. From the life-or-death evaluations of academy research institutes in the post-communist countries to the Government Performance and Results Act in the United States, research evaluation has become a topic of utmost importance in science policy. Because evaluation often have substantial consequences for researchers and research institutions, including restructuring, shifting of priorities, budget reductions, or even closures, it is essential that they be done systematically and objectively, with methodologies that can be understood and trusted by those concerned. This book, based on a NATO Advanced Research. Workshop co-organised by the Academy of Sciences of the Czech Republic and the American Association for the Advancement of Science, describes a range of the most up-to-date methods of science evaluation and the experience in their implementation in many countries. It will be of interest to researchers, policy-makers, practitioners of science evaluation, and many others
interested in science policy.
http://www.iospress.nl/html/stp.html

20. Association of European Universities CRE

Hochschulevaluation

Barblan, Andris (1997). Management for Quality: The CRE Programme of Institutional Evaluation - Issues Encountered in the Pilot Phase, 1994/95. In John Brennan, Peter de Vries & Ruth Williams (Eds.), Standards and Quality in Higher Education (174-197). London: Jessica Kingsley Publishers.

The Association of European Rectors (CRE) aims to develop an external supportive review system. The chapter by Barblan describes three pilot evaluations of the universities of Göteborg (Sweden), Oporto (Portugal), and Utrecht (The Netherlands). A three person review team visited each institution and, in the main site visit to each university, met with over 80 'witnesses' over three days. The focus was upon external constraints and institutional norms, and their influence upon the universities' capacity for change. The CRE project is largely about the contribution which external quality assessment can make to the management of institutional change in higher education.

Barblan, Andris (1996). Institutional evaluation: Assessing the pilot phase. CRE-Action 107, 55-73.

Vught, Frans A. van & Westerheijden, D. F. (1996). Institutional evaluation and management for quality - The CRE programme: Background, goals and procedures. CRE-Action 107, 9-40.

21. UNESCO

Hochschulevaluation

Cowen, Robert (Ed.). (1996). World Yearbook of Education 1996: The Evaluation of Higher Education Systems. London: Kogan Page.

The authors of the book deal with the following topics: (1) traditions of academic excellence in higher education; (2) the emergence of ideologies of evaluation for higher education; (3) the social, political and economic pressures which have produced higher education evaluation systems; (4) national styles of higher education evaluation; and (5) the consequences of evaluation for university funding, teaching and research, student opportunities and communication between the state and its higher education system.

UNESCO (Ed.). (1995). Quality Assurance and Institutional Accreditation in European Higher Education. Higher Education in Europe, 20(1-2), 1-201.

Vessuri, Hebe (Ed.). (1993). La Evaluación Academica. Paris: UNESCO-CRE.